

Host preference explains the high endemism of ectomycorrhizal fungi in a dipterocarp rainforest

Hirotooshi Sato¹  | Ajuwin Lain² | Takafumi Mizuno¹ | Satoshi Yamashita³ |
Jamilah Binti Hassan² | Khairunnisa Binti Othman⁴ | Takao Itioka¹

¹Graduate School of Human and Environmental Studies, Kyoto University, Sakyo, Kyoto, Japan

²Sarawak Biodiversity Centre, Kuching, Sarawak, Malaysia

³Center for Biodiversity and Climate Change, Forestry and Forest Products Research Institute (FFPRI), Tsukuba, Ibaraki, Japan

⁴Research and Development Division, Forest Department Sarawak, Kuching, Sarawak, Malaysia

Correspondence

Hirotooshi Sato, Graduate School of Human and Environmental Studies, Kyoto University, Sakyo, Kyoto 606-8501, Japan.
Email: h-sato@sys.bot.kyoto-u.ac.jp

Funding information

Science and Technology Research Partnership for Sustainable Development, Grant/Award Number: JPMJSA190; Japan Society for the Promotion of Science, Grant/Award Number: 20K06796

Handling Editor: Isheng J. Tsai

Abstract

Ectomycorrhizal (ECM) fungi are important tree symbionts within forests. The biogeography of ECM fungi remains to be investigated because it is challenging to observe and identify species. Because most ECM plant taxa have a Holarctic distribution, it is difficult to evaluate the extent to which host preference restricts the global distribution of ECM fungi. To address this issue, we aimed to assess whether host preference enhances the endemism of ECM fungi that inhabit dipterocarp rainforests. Highly similar sequences of 175 operational taxonomic units (OTUs) for ECM fungi that were obtained from Lambir Hill's National Park, Sarawak, Malaysia, were searched for in a nucleotide sequence database. Using a two-step binomial model, the probability of presence for the query OTUs and the registration rate of barcode sequences in each country were simultaneously estimated. The results revealed that the probability of presence in the respective countries increased with increasing species richness of Dipterocarpaceae and decreasing geographical distance from the study site (i.e. Lambir). Furthermore, most of the ECM fungi were shown to be endemic to Malaysia and neighbouring countries. These findings suggest that not only dispersal limitation but also host preference are responsible for the high endemism of ECM fungi in dipterocarp rainforests. Moreover, host preference likely determines the areas where ECM fungi potentially expand and dispersal limitation creates distance-decay patterns within suitable habitats. Although host preference has received less attention than dispersal limitation, our findings support that host preference has a profound influence on the global distribution of ECM fungi.

KEYWORDS

biodiversity, dispersal limitation, DNA barcoding, host specificity, mycorrhizal symbiosis, tropical rainforest

1 | INTRODUCTION

Although the biogeographic realms of terrestrial animals and plants have been well-investigated (Holt et al., 2013; Kreft & Jetz, 2010; Morrone, 2018), the development of biogeographic studies on fungi has lagged. This is primarily because of the paucity of morphological characters in fungi, which can lead to confusion between multiple species that are morphologically similar (Balasundaram et al., 2015; Bickford et al., 2007). Moreover, there is often insufficient data available on the large-scale distribution of the respective fungal species, including mushroom-forming fungi (Peay & Matheny, 2016; Sato et al., 2012). However, the recent development of DNA-based approaches has contributed to determining fungal biogeographic patterns (Duarte et al., 2016; Peay & Matheny, 2016; Tedersoo, 2017; Werth, 2011). Sequencing of the rRNA internal transcribed spacer (ITS) regions, that is, the barcode locus for fungi, has become an effective tool for distinguishing between cryptic fungal species (Köljalg et al., 2013; Nilsson et al., 2018; Schoch et al., 2012; Smith et al., 2007; Větrovský et al., 2020), making it easy to assess whether two fungal collections are conspecific. In addition, the DNA metabarcoding technologies have allowed for the collection of distribution data from environmental samples, such as soil and plant tissues (Köljalg et al., 2013; Peay & Matheny, 2016; Schmidt et al., 2013). Although an increased number of studies have examined the biogeography of fungi in the past few decades, fungal biogeography at a large scale still remains to be elucidated (Duarte et al., 2016; Peay & Matheny, 2016; Tedersoo, 2017; Werth, 2011).

Recent studies focusing on fungal biogeography have presented evidence that fungal species are not necessarily cosmopolitan. The classic Baas-Becking hypothesis, 'Everything is everywhere, but the environment selects' suggests that microbial taxa are found anywhere on earth where there is suitable habitat for them (De Wit & Bouvier, 2006). This is, to an extent, true for some fungal guilds, such as arbuscular mycorrhizal fungi (Davison et al., 2015; Stürmer et al., 2018), ericoid mycorrhizal fungi (Kohout, 2017), root-endophytic fungi (Glynou et al., 2016; Queloz et al., 2011) and soil-borne saprotrophic fungi (Egidi et al., 2019; Tedersoo et al., 2014), for which inter-continental distribution is often observed. However, fungal taxa with an almost cosmopolitan distribution are less common than traditionally thought. For instance, studies using DNA metabarcoding technologies have indicated that most soil-inhabiting fungal taxa are not cosmopolitan but are rather endemic to particular bioregions (Talbot et al., 2014; Tedersoo et al., 2014, 2022; Větrovský et al., 2019).

Ectomycorrhizal (ECM) fungi are one of the most well-known soil-inhabiting fungi that are believed to have restricted distribution ranges (Peay, Garbelotto, & Bruns, 2010; Peay & Matheny, 2016; Sato et al., 2012; Talbot et al., 2014; Tedersoo et al., 2014, 2022; van der Linde et al., 2018). ECM fungi are ubiquitous mutualistic symbionts of trees that are dominantly found in temperate and tropical forests, such as those belonging to Pinaceae, Fagaceae, Betulaceae, Dipterocarpaceae and Myrtaceae (Smith & Read, 2008;

Tedersoo et al., 2010; Tedersoo & Smith, 2013). There are approximately 20,000–25,000 (Rinaldi et al., 2008) or more (Sato, 2024) ECM fungal species on earth, which are known to decline moderately towards lower latitudes (Mikryukov et al., 2023; Tedersoo & Nara, 2010). Although ECM symbiosis is almost ubiquitous in forest ecosystems, previous studies have shown a strong distance–decay relationship in ECM fungal communities at both large and local scales (Bahram et al., 2013; Bogar & Peay, 2017; Peay, Garbelotto, & Bruns, 2010; Sato et al., 2012; Talbot et al., 2014; Tedersoo et al., 2014, 2022).

An important question in the field of ECM fungal biogeography is to what extent host–fungus compatibility controls the biogeography of ECM fungi on a global scale. Distribution ranges of ECM fungal species are restricted by dispersal limitations that can be caused by geographical barriers (e.g. mountain and ocean) and the limited dispersal ability of microscopic propagules (Bahram et al., 2013; Bogar & Peay, 2017; Glassman et al., 2015; Peay et al., 2012; Peay, Garbelotto, & Bruns, 2010). It is worth noting that ECM fungi exhibit stronger distance–decay patterns than saprotrophic fungi (Sato et al., 2012). A potentially important factor to explain this is the compatibility of ECM fungi with host plants, that is, host specificity and host preference. For instance, the distribution of host plants was highly correlated with that of ECM fungi at larger scales (Delhaye et al., 2024; Pölme et al., 2013; Sato et al., 2012; van der Linde et al., 2018). However, caution should be exercised when interpreting the effect of host distribution because there are not many host-specific ECM fungi (Molina et al., 1992; Molina & Horton, 2015; Tedersoo et al., 2008) (but also see Tedersoo et al., 2024). Furthermore, the Holarctic distribution is so common in ECM plant families (e.g. Pinaceae, Fagaceae and Betulaceae) and genera (e.g. *Abies*, *Pinus*, *Quercus*, *Betula* and *Alnus*) (Heywood, 1978) that difficulty is encountered in testing whether host specificity or preference for either of these plant taxa could control the ECM fungal biogeography at a global scale. Therefore, an important challenge in ECM fungal biogeography is to evaluate the influence of host distribution on a global scale more confidentially.

This study aimed to test the hypothesis that host distribution is the major determinant for controlling the distribution range of ECM fungi on a global scale. To determine this, we estimated the global distribution of ECM fungi that were found in a dipterocarp rainforest in Lambir Hill's National Park, Sarawak, Malaysia, using ITS barcodes obtained from international nucleotide sequence databases. We focused on ECM fungi in a dipterocarp rainforest because the biogeography of Dipterocarpaceae does not overlap with that of other ECM plant taxa, which allows for the confidential evaluation of host effects on ECM fungal biogeography on a global scale. In this study, we used a modified version of the two-step binomial mode that was previously developed (Sato et al., 2012) to address the incomplete and biased registration rates of ITS barcodes among the different countries. Based on the estimation by modelling, we evaluated to what extent the distribution range of ECM fungi in the study site (i.e. Lambir Hill's National Park) could be explained by variables such as hosts, geographical distance and climatic variables.

2 | MATERIALS AND METHODS

2.1 | Dataset

In this study, 175 ECM fungal operational taxonomic units (OTUs) that were found in Lambir Hill's National Park, Sarawak, Malaysia (N4°12', E114°02', 130–150 m a.s.l.), were used for the analysis (hereafter, referred to 'query OTUs'). As Lambir Hill's National Park is known as one of the most tree-diverse forests in the Palaeotropical region (Ashton, 2005), we selected this area as the study site for estimating the global distribution of ECM fungi inhabiting dipterocarp rainforests. Three datasets were prepared for this study, including Dataset 1 (108 OTUs detected in Sato et al. (2015)), Dataset 2 (89 OTUs detected in Peay, Kennedy, et al. (2010)) and Dataset 3 (175 OTUs detected in these two studies). These OTUs were collected from ECM root samples. All query OTUs used in the analysis are available in the Dryad repository (<https://doi.org/10.5061/dryad.k3j9kd5gs>). The study area consists of approximately 6500 ha of primarily lowland, and mixed tropical forest, which is dominated by Dipterocarpaceae. The primary and secondary ECM hosts in the area are Dipterocarpaceae and Fagaceae (especially *Lithocarpus*), although the latter trees are relatively rare. The rocks are sedimentary and are composed of alternating layers of sandstone and shale.

2.2 | BLAST search

Fungal materials that showed highly similar sequences with the query OTUs were searched at the global scale using the basic local alignment search tool (BLAST) available on the National Center for Biotechnology Information (NCBI), UNITE (<https://unite.ut.ee/>) and GlobalFungi (<https://globalfungi.com/>) sites. The target fungal materials in the searches were restricted to the fruitbody and ECM root samples to ensure that focal fungal species were established in areas where they were detected. The geographical origins of the highly similar hit sequences (i.e. $\geq 97\%$ sequence identity and $\geq 70\%$ query coverage with query sequences) were recorded. In addition, 98% sequence similarity was used to ensure that the results were robust for the threshold of sequence similarity between the query and hit sequences. The number of query OTUs for which highly similar hit sequences were found were then tallied for each country, except for the five largest countries in the world (Russia, Canada, China, United States and Brazil), which were further divided into five to eight regions each (Table S1), and the number of query OTUs was counted at the regional level.

2.3 | Modelling

The modelling used in this study aimed to reveal how many ECM fungal OTUs in the study site are shared with respective countries/regions. This model postulates that failure to detect ECM

fungi in a particular country can occur because either the focal ECM fungi are absent or they are present, but their ITS barcodes have not yet been registered in the evaluated database. To address this, the observed data in this study were fitted to a modified version of the two-step binomial model that was used in a previous study (Sato et al., 2012). In this modelling, the number of OTUs detected in a particular country/region is drawn from two binomial processes. The first binomial process controls the probability that an arbitrary fungal OTU in the study site is present in country/region i in Dataset j (p_{ij}). This process also comprises the number of fungal OTUs that exist in country/region i in Dataset j (E_{ij}) among the total number of fungal OTUs detected in the study site in Dataset j (N_j). The second binomial process controls the proportion of fungal OTUs whose ITS barcodes have already been registered in the nucleotide sequence database in country/region i in Dataset j (r_{ij}). This process also comprises the number of fungal OTUs that are detected in country/region i in Dataset j (D_{ij}) among E_{ij} . Among these parameters, N_j and D_{ij} are known parameters and the remaining are unknown parameters to be estimated. Ignoring the influence of covariates, the two-step binomial model is drawn as follows:

$$\Pr[D_{ij} = 0] = (1 - p_{ij})^{N_j} + \binom{N_j}{E_{ij}} p_{ij}^{E_{ij}} (1 - p_{ij})^{N_j - E_{ij}} \times (1 - r_{ij})^{E_{ij}}$$

$$\Pr[D_{ij} > 0] = \binom{N_j}{E_{ij}} p_{ij}^{E_{ij}} (1 - p_{ij})^{N_j - E_{ij}} \times \binom{E_{ij}}{D_{ij}} r_{ij}^{D_{ij}} (1 - r_{ij})^{E_{ij} - D_{ij}}$$

where each parameter is defined above.

Multiple logistic regression was used to estimate the probability of presence (p_{ij}) based on several explanatory variables, including the species richness of primary (S1) and secondary (S2) host plants, geographical distance from the study site (GD), annual mean temperature (AT) and annual accumulative precipitation (AP) in each country/region (Table S2). Here, the primary and secondary host plants correspond to Dipterocarpaceae and *Lithocarpus* (Fagaceae), respectively, because they are the most dominant and second most dominant ECM trees in the study site, respectively. The global distribution of respective species belonging to Dipterocarpaceae and *Lithocarpus* was examined from the Plants of the World Online database hosted by the Royal Botanic Garden Kew (<https://powo.science.kew.org/>). The species richness of the primary and secondary hosts was used as proxies for the predominance of these hosts. Geographical distance was calculated based on the latitude and longitude of the study site and the capital city in each country/region. A set of bioclimatic variables for each country/region, including annual temperature (bio1) and precipitation (bio12), was downloaded from WorldClim v2.1 (<http://www.worldclim.org>) at a resolution of 2.5 arc minutes using the 'worldclim_global' function of the 'geodata ver. 0.5.8' R packages (Hijmans et al., 2023). The resulting layer was converted into a data frame using the 'rasterToPoints' function of the 'raster ver. 3.6.23' R package (Hijmans et al., 2015). The two bioclimatic variables

were then averaged over all data points that were located within a 3.0×3.0min grid centred on the capital of each country/region. Therefore, p_i is described as follows:

$$\text{logit}(p_{ij}) = \alpha_j + \sum_{k=1}^K \beta_{jk} x_k$$

where logit is the logistic transformation, x_k is the value of the k th explanatory variable, α_j is an intercept and β_{jk} is the partial regression coefficient estimated for the k th explanatory variable in Dataset j . Among the five explanatory variables, the log transformation (i.e. $\log(x_i + 1)$) was applied to S1, S2 and GD to reduce the skewness. Except for GD, two different schemes were prepared for the explanatory variables; the values of the respective variables obtained in the focal country/region were directly used for the first scheme (raw), whereas the absolute difference of the values between the study site and focal country/region was used for the second scheme (abs. dif.). A non-informative prior distribution was used to estimate α_j and β_{jk} . In addition, the multicollinearity was pre-checked based on the tolerance value and variance inflation factor (VIF), where a VIF value ≥ 10 is often used as an indication of a potential multicollinearity problem.

The deposition rate of ITS barcodes in each country/region in Dataset j (r_{ij}) was modelled using the Beta distribution. The Beta distribution can be used to estimate the posterior for the unknown rate, in which two parameters a and b are usually interpreted as the prior number of successes and failures respectively (Gupta & Nadarajah, 2004). To assess prior successes and failures in our model, there was a need to clarify how many ECM fungal OTUs available in the NCBI were non-singleton (OTUs occurring more than once) or singleton (OTUs occurring only once) because non-singleton and singleton OTUs are considered as prior successes and failures, respectively, based on leave-one-out cross validation. At first, the 88,354 ITS barcodes of ECM fungal taxa were downloaded from the NCBI website. The resulting ITS barcodes were clustered into 15,479 OTUs at the 97% sequence similarity threshold using the 'cldasseq' command in Claident ver. 0.2.2019.05.10 (Tanabe & Toju, 2013). Then, the number of non-singleton and singleton OTUs was counted for each country/region. Therefore, r_{ij} is described as follows:

$$r_{ij} \sim \text{Beta}(a_i + 1, b_i + 1)$$

where a_i and b_i are the number of non-singleton and singleton OTUs in country/region i respectively. Notably, this prior is almost equivalent to a non-informative prior when the total number of ECM fungal OTUs in the focal country/region is zero or nearly zero. To reduce the spurious estimation, two schemes were prepared to estimate the deposition rates; the countries where the total number of ECM fungal OTUs was < 5 and 1 were precluded from the analysis in the first and second schemes respectively (i.e. 102 and 127 countries in the first and second schemes respectively). Overall, there were 24 schemes present in this modelling, including three datasets (Datasets 1, 2 and 3), two settings for sequence similarity threshold

in BLAST search (i.e. 97% and 98%), two schemes for setting explanatory variables (i.e. raw value and absolute difference) and two schemes for the number of focal countries in the estimation (i.e. 102 and 127 countries).

The analyses were run in a Bayesian framework using the Stan probabilistic programming language, in R ver. 4.3.0 (R Core Team, 2023) using the 'rstan ver. 2.26.23' package (Stan Development Team, 2023). The default Hamilton Monte Carlo approach was used as a model-fitting algorithm within rstan. The models were run with four chains of 5000 iterations and thinned every 5th iteration. The first 1000 iterations were burn-in iterations to guarantee convergence to the posterior distributions (resulting in a total of $800 \times 4 = 3200$ samples). The convergence of the four chains was tested using the Gelman–Rubin convergence diagnostic statistic (i.e. Rhat < 1.1) (Gelman & Rubin, 1992). In addition, an effective sample size of each parameter was confirmed to be ≥ 200 to ensure sufficient sampling. The Stan code and R commands are shown in the Dryad repository (<https://doi.org/10.5061/dryad.k3j9kd5gs>).

Subsequent analyses were performed using R ver. 4.3.0. A posterior density of the standard partial regression coefficient for each explanatory variable was plotted using the 'ggplot2 ver.3.4.3' and 'ggsci ver. 3.0.0' R packages after reshaping the values obtained in rstan using the 'melt' function of the 'reshape2 ver. 1.4.4' R package. In the multiple logistic regression for the full model, a one-tailed test was used to confirm whether the partial regression coefficient of each explanatory variable was significantly positive or negative. Multiple testing problems were addressed using the false discovery rate (FDR) correction that was applied to the p-values obtained for each explanatory variable. The world map was obtained using the 'map_data' function of the 'maps ver.3.4.1' R package. The probability of occurrence (p_i) and the ITS deposition rate (r_i) were then plotted on the world map using the ggplot2 and 'tidyverse ver. 2.0.0' R packages. The leave-one-out cross-validation information criterion (LOOIC) (Vehtari et al., 2017; Watanabe 2010) was used to calculate the expected log predictive density (ELPD) and to select the most appropriate model from the full and reduced models in the logistic regression of p_i (i.e. the model with the lowest LOOIC was the most appropriate). If the difference in the ELPD (elpd_diff) between the models was < 4 , the difference between the models was considered not important. The LOOIC was calculated using the 'loo' function of the 'loo ver. 2.6.0' R package (Vehtari et al., 2021). In total, 16 models were compared, including M1 (full model; raw), M1d (full model; abs. dif.), M2 (S1 + GD + AT + AP; raw), M2d (S1 + GD + AT + AP; abs. dif.), M3 (S1 + S2 + GD + AP; raw), M3d (S1 + S2 + GD + AP; abs. dif.), M4 (S1 + S2 + AT + AP; raw), M4d (S1 + S2 + AT + AP; abs. dif.), M5 (S1 + S2 + GD + AT; raw), M5d (S1 + S2 + GD + AT; abs. dif.), M6 (S2 + GD + AT + AP; raw), M6d (S2 + GD + AT + AP; abs. dif.), M7 (S1 + GD + AP; raw), M7d (S1 + GD + AP; abs. dif.), M8 (S1 + GD; raw), and M8d (S1 + GD; abs. dif.). Furthermore, a linear correlation between each explanatory variable and the posterior mean for the probability of presence (p_{ij}) that was estimated using the full model was plotted using the 'geom_point' and 'geom_smooth' functions of the ggplot2 R package.

3 | RESULTS

3.1 | Detection of highly similar sequences

The number of hit sequences with $\geq 97\%$ sequence identity in Dataset 1, 2 and 3 was 64, 33 and 96 respectively (Table S3). These hit sequences were available from both the NCBI (90 accessions) and UNITE (6 accessions) databases. No highly similar hit sequences were found from the Global Fungi database, presumably due to the low sequence coverage of the database sequences. When the threshold of sequence identity between the query and hit sequences was set to $\geq 97\%$, the numbers of query OTUs for which highly similar hit sequences were obtained were 38, 39 and 55 in Dataset 1, 2 and 3 respectively (Table S3). These numbers decreased slightly when the threshold of sequence identity was set to $\geq 98\%$, which resulted in 35, 36 and 51 query OTUs in Dataset 1, 2 and 3 respectively. The most predominant geographic origin of highly similar sequences was Malaysia, including both Peninsula Malaysia and Borneo; the database sequences originating from Malaysia showed a $\geq 97\%$ sequence identity with 33, 33 and 44 query OTUs in Dataset 1, 2 and 3 respectively (Tables 1 and S3). The second most predominant geographic origin was Indonesia, including Jawa, Sumatra and Bangka Island; sequences originating from Indonesia showed a $\geq 97\%$ sequence identity with 9, 8 and 14 query OTUs in Dataset 1, 2 and 3 respectively (Tables 1 and S3). Several query OTUs showed a $\geq 97\%$ sequence identity with the database sequences originating from tropical areas in Southeast and South Asia, such as Papua New Guinea (2, 0 and 2 query OTUs in Dataset 1, 2 and 3 respectively), Singapore (0, 1 and 1 query OTUs in Dataset 1, 2, and 3 respectively), Thailand (0, 3 and 3 query OTUs in Dataset 1, 2 and 3 respectively), Laos (1, 0 and 1 query OTUs in Dataset 1, 2 and 3 respectively), Bangladesh (0, 1

TABLE 1 Results of the global search for sequences that are highly similar to the ectomycorrhizal (ECM) fungal operational taxonomic units (OTUs) detected in Lambir Hill's National Park using the basic local alignment search tool (BLAST).

Region	Dataset 1 (N=108)	Dataset 2 (N=89)	Dataset 3 (N=175)
Malaysia	33 (29)	33 (32)	44 (40)
Indonesia	9 (8)	8 (8)	14 (13)
Papua New Guinea	2 (2)	0 (0)	2 (2)
Singapore	0 (0)	1 (1)	1 (1)
Thailand	0 (0)	3 (3)	3 (3)
Laos	1 (1)	0 (0)	1 (1)
Bangladesh	0 (0)	1 (1)	1 (1)
Sri Lanka	0 (0)	1 (0)	1 (0)
China Southeast	2 (2)	2 (2)	3 (3)
Japan	1 (1)	1 (1)	1 (1)

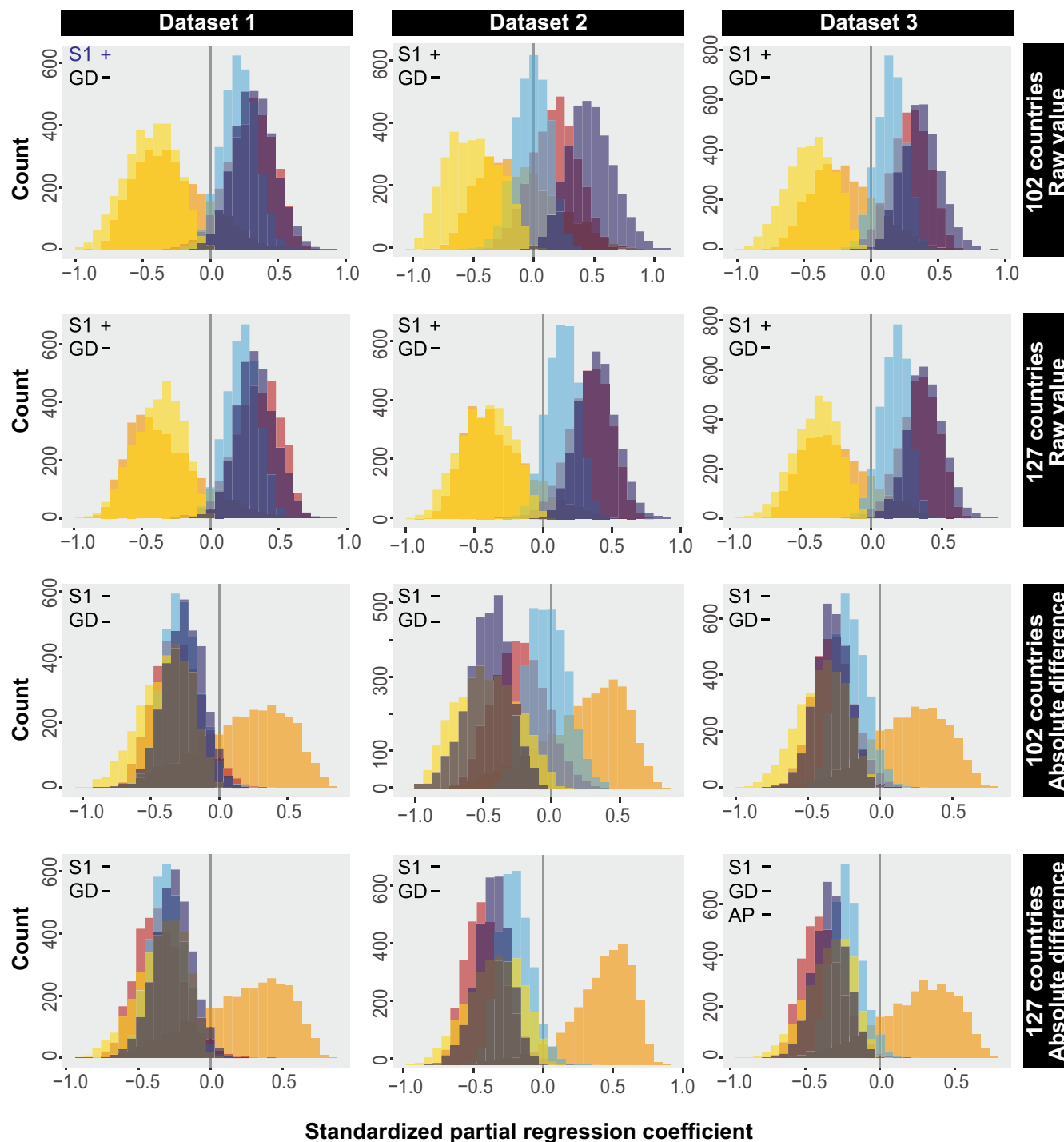
Note: The number in each cell represents the number of query OTUs for which hit sequences with $>97\%$ (or $>98\%$) sequence identity were found in a focal country. N represents the total number of query OTUs in each dataset.

and 1 query OTUs in Dataset 1, 2 and 3 respectively) and Sri Lanka (0, 1 and 1 query OTUs in Dataset 1, 2 and 3 respectively) (Tables 1 and S3). In addition, hit sequences with a $\geq 97\%$ sequence identity were partially found in the subtropical areas in East Asia, such as China (2, 2 and 3 query OTUs in Dataset 1, 2 and 3 respectively; the geographical origins were Hunan, Jiangxi and Guangdong provinces) and Japan (1, 1 and 1 query OTUs in Dataset 1, 2 and 3 respectively; geographical origin was Yakushima Island in Kyushu) (Tables 1 and S3). When the threshold of the sequence identity was set to $\geq 98\%$, the results changed in several countries. The number of query OTUs that were highly similar to the hit sequences found in the focal countries decreased in Malaysia (29, 32 and 40 query OTUs in Dataset 1, 2 and 3 respectively), Indonesia (8, 8 and 13 query OTUs in Dataset 1, 2 and 3 respectively) and Sri Lanka (0, 0 and 0 query OTUs in Dataset 1, 2 and 3 respectively).

3.2 | Results of modelling

Results of the multiple logistic regression showed that S1 and GD had significant and strongest effects on the probability of presence of ECM fungal OTUs in the study site (Figures 1 and S1). In all the results, the standardized partial regression coefficients for S1 were significantly positive and negative when the raw and abs. dif. were used as the explanatory variables, respectively (Figures 1 and S1). Moreover, the standardized partial regression coefficients for GD were significantly negative, except for one setting (Dataset=1, similarity threshold=98%, country=102, variable type=abs. dif.) (Figures 1 and S1). The absolute values of the standardized partial regression coefficients for these two variables were substantially high in all results. The standardized partial regression coefficients for S2 and AT were neither significantly positive nor negative in all the results. The standardized partial regression coefficients for AP were significantly positive in one setting (Dataset=1, similarity threshold=98%, country=102 and variable type=raw) and significantly negative in two settings (Dataset=1, similarity threshold=97%, country=102 and variable type=abs. dif.; Dataset=3, similarity threshold=97%, country=127 and variable type=abs. dif.). A correlation between the probability of presence and each explanatory variable also revealed that the probability of presence changed substantially depending on S1 and GD but changed only slightly with regard to AT and AP (Figures S2–S4). Overall, S1 and GD had a strong influence on the probability of presence, regardless of the model settings, whereas the influence of the other factors was not necessarily supported by the data.

The results of the model selection were almost similar to those of the standardized partial regression coefficients. The models with the lowest LOOIC for the first (Dataset 1 and 102 countries), second (Dataset 2 and 102 countries), third (Dataset 3 and 102 countries), fourth (Dataset 1 and 127 countries), fifth (Dataset 2 and 127 countries) and sixth (Dataset 3 and 127 countries) settings were M8d, M8d, M3d, M8d, M2 and M3d respectively (Table 2). However, in total, 13, 12, 11, 12, 14 and 12 models were shown to be equally



Explanatory variable

- Species richness of primary host (S1)
- Species richness of secondary host (S2)
- Geographical distance (GD)
- Annual mean temperature (AT)
- Annual accumulative precipitation (AP)

+ Positive effect (>95% posterior probability)

- Negative effect (>95% posterior probability)

supported ($\text{elpd_diff} < 4$) under the first, second, third, fourth, fifth and sixth settings respectively (Table 2). All the equally supported models included S1 as an explanatory variable, regardless of the

model settings (Table 2). In the second, third and sixth settings, all the equally supported models included GD as an explanatory variable (Table 2). In contrast, the models in which S2, AT and AP were

FIGURE 1 Posterior distribution of the standardized partial regression coefficient calculated in multiple logistic regression modelling the probability of presence of ectomycorrhizal fungal operational taxonomic units (OTUs) detected in Lambir Hill's National Park using a 97% threshold of sequence identity between the query OTUs and hit sequences in the database. The absolute value of the standardized partial regression coefficient represents the impact of each explanatory variable on the probability of presence. Positive and negative values indicate positive and negative effects, respectively. The results are shown according to the difference in the datasets used in the estimation (Dataset 1, 2 or 3), the setting of explanatory variables (i.e. raw value and absolute difference between Lambir Hill's National Park and the focal country/region), and the number of focal countries in the estimation (i.e. 102 and 127 countries). Primary and secondary hosts correspond to Dipterocarpaceae and *Lithocarpus*, respectively. The posterior distributions for the different explanatory variables are shown in different colours. An explanatory variable that shows a positive (+) or negative (−) effect with a >95% posterior probability (the false discovery rate (FDR) correction was used) is shown in each graph.

removed (M8 and M8d) had sufficiently low LOOIC values (elpd_diff <4) under all the settings (Table 2), indicating that S2, AT and AP were not especially important for predicting the probability of presence.

The probability of presence was particularly high in the neighbouring countries of Malaysia, in which dipterocarp trees are present (Figure 2, Table S4). Except for Malaysia, the probability of presence in the most appropriate models was highest in Indonesia, where the mean probabilities ($\pm$ standard deviation (SD)) in the first, second, third, fourth, fifth and sixth settings were 23.4% ($\pm 8.3\%$), 23.8% ($\pm 8.5\%$), 42.1% ($\pm 13.7\%$), 47.2% ($\pm 15.4\%$), 36.0% ($\pm 11.1\%$) and 37.9% ($\pm 11.4\%$) respectively (Figure 2, Table S4). This was followed by the Philippines (7.6% ($\pm 3.4\%$), 6.5% ($\pm 2.7\%$), 8.2% ($\pm 3.7\%$), 13.9% ($\pm 6.8\%$), 9.7% ($\pm 3.6\%$) and 9.2% ($\pm 3.2\%$)), Singapore (8.6% ($\pm 5.6\%$), 6.8% ($\pm 4.2\%$), 7.9% ($\pm 5.4\%$), 15.5% ($\pm 9.8\%$), 8.6% ($\pm 4.8\%$) and 7.5% ($\pm 3.7\%$)), Thailand (6.2% ($\pm 1.8\%$), 6.0% ($\pm 1.8\%$), 8.6% ($\pm 2.7\%$), 5.1% ($\pm 2.5\%$), 5.1% ($\pm 1.7\%$) and 4.7% ($\pm 1.6\%$)) and Vietnam (2.8% ($\pm 0.8\%$), 2.6% ($\pm 0.8\%$), 2.8% ($\pm 1.0\%$), 3.9% ($\pm 1.3\%$), 4.1% ($\pm 1.2\%$) and 4.1% ($\pm 1.2\%$)) respectively (Figure 2, Table S4). Among the countries where dipterocarp trees are absent, the probability of presence in the first, second, third, fourth, fifth and sixth settings was highest in Taiwan (0.2% ($\pm 0.3\%$), 0.1% ($\pm 0.2\%$), 0.1% ($\pm 0.1\%$), 0.5% ($\pm 1.0\%$), 1.0% ($\pm 1.2\%$) and 1.1% ($\pm 1.4\%$)). The probability of presence was almost similar across the equally supported models under all the settings, except for Indonesia. In Indonesia, the mean probabilities were remarkably high in the models where GD was removed (M4 and M4d), ranging from 61.3% (Dataset 1 and 127 countries) to 79.7% (Dataset 2 and 127 countries), whereas they were relatively low in models where S2, AT, and AP were removed (M8 and M8d), ranging from 23.4% (Dataset 1 and 102 countries) to 42.1% (Dataset 2 and 102 countries) (Table S4).

As expected, the deposition rate of the ITS barcodes in the database was higher in temperate areas than in tropical areas. Regardless of the model settings, the deposition rate was especially high in Hungary, New Zealand, the United States, Canada, and China (Table S4). Although the SD was high, the mean deposition rate was also high in Bangladesh and Greenland (Table S4). When the number of focal countries was 127, the deposition rates were particularly high in Venezuela, Azerbaijan, and Armenia, but these rates were likely overestimated (Table S4). Except for Thailand, the deposition rate was relatively low in Southeast Asian countries, such

as Malaysia, Cambodia, Myanmar, Singapore, Indonesia, and the Philippines (Table S4).

4 | DISCUSSION

These findings have important implications for the understanding of what determines the ECM fungal biogeography on a global scale. In our modelling, the distribution of Dipterocarpaceae trees and the geographical distances were shown as the primary determinants for the global distribution of ECM fungi in the study site. This result suggests that not only dispersal limitation but also host–fungus compatibility strongly control the distribution ranges of ECM fungi on a global scale, which results in high endemism of ECM fungi in dipterocarp rainforests, although caution should be exercised for the limitations of the modelling.

Our results support the hypothesis that host–fungus compatibility can control the areas where ECM fungi can expand on a global scale. The modelling results showed that the abundance of Dipterocarpaceae had a strong influence on the probability of occurrence (Figures 1 and S1, Table 2), suggesting that the vast majority of ECM fungi in the study site cannot expand into areas where dipterocarp trees are absent. These findings indicate that host–fungus compatibility is one of the most important determinants for ECM fungal biogeography at a large scale, which is in agreement with previous studies (Delhaye et al., 2024; Pölme et al., 2013; Sato et al., 2012; van der Linde et al., 2018). However, this view is counterintuitive given that strict host specificity phenomena do not appear to be common in ECM fungi (Molina et al., 1992; Molina & Horton, 2015; Tedersoo et al., 2008). This is likely true for ECM fungi in dipterocarp rainforests (Peay et al., 2015; Sato et al., 2015). Nevertheless, many ECM fungi appear to exhibit host preference and an association with a preferred specific host genus or family (Delhaye et al., 2024; Ishida et al., 2007; Tedersoo et al., 2008, 2009, 2024; van der Linde et al., 2018), suggesting that host preference, rather than host specificity, can restrict the distribution ranges of ECM fungi. Notably, ECM fungi that preexist in forests usually have an advantage over later immigrants in earlier access to root tips and subsequent ECM colonization (i.e. priority effect) (Kennedy, 2010; Kennedy et al., 2009), which may increase the chance that preexisting ECM fungi are associated with non-preferable hosts. In other words, immigrant ECM fungi may be more stringently dependent on preferable hosts than preexisting ECM fungi. Therefore, the

TABLE 2 Comparisons between the multiple logistic regression modelling of the probability of presence of the ectomycorrhizal fungal operational taxonomic units (OTUs) detected in Lambir Hill's National Park based on the leave-one-out cross-validation information criterion (LOOIC).

Model	Explanatory variables					Dataset 1			Dataset 2			Dataset 3		
	S1	S2	GD	AT	AP	Type	102 Countries	127 Countries	102 Countries	127 Countries	102 Countries	127 Countries	102 Countries	127 Countries
M1	○	○	○	○	○	Raw	-1.6 (4.7)	-1.6 (4.9)	-3.8 (2.0)	-2.2 (1.9)	-3.0 (1.5)	-2.2 (1.9)	-3.0 (1.5)	-0.8 (1.6)
M1d	○	○	○	○	○	Abs. Dif.	-2.6 (5.1)	-2.2 (5.0)	-3.2 (1.7)	-0.4 (2.0)	-1.1 (1.4)	-0.4 (2.0)	-1.1 (1.4)	-0.7 (1.5)
M2	○	○	○	○	○	Raw	-3.1 (2.5)	-2.0 (3.0)	-1.9 (1.7)	0.0 (0.0)	-1.6 (1.9)	0.0 (0.0)	-1.6 (1.9)	-0.6 (2.2)
M2d	○	○	○	○	○	Abs. Dif.	-3.4 (2.2)	-3.7 (2.5)	-1.5 (1.5)	-0.4 (0.7)	-1.4 (2.1)	-0.4 (0.7)	-1.4 (2.1)	-1.3 (2.6)
M3	○	○	○	○	○	Raw	-1.3 (2.7)	-2.0 (3.2)	-2.0 (1.3)	-2.1 (1.7)	-1.6 (1.1)	-2.1 (1.7)	-1.6 (1.1)	-1.2 (1.3)
M3d	○	○	○	○	○	Abs. Dif.	-1.5 (3.3)	-2.0 (3.6)	-1.9 (1.1)	-1.3 (1.8)	0.0 (0.0)	-1.3 (1.8)	0.0 (0.0)	0.0 (0.0)
M4	○	○	○	○	○	Raw	-3.6 (9.6)	-3.7 (9.1)	-6.9 (5.7)	-3.7 (4.7)	-8.8 (7.0)	-3.7 (4.7)	-8.8 (7.0)	-7.3 (6.1)
M4d	○	○	○	○	○	Abs. Dif.	-3.7 (9.8)	-3.7 (9.1)	-4.9 (4.9)	-0.3 (3.7)	-5.9 (5.8)	-0.3 (3.7)	-5.9 (5.8)	-4.1 (5.0)
M5	○	○	○	○	○	Raw	-2.7 (2.6)	-4.2 (3.4)	-2.3 (1.2)	-2.7 (1.6)	-3.5 (2.4)	-2.7 (1.6)	-3.5 (2.4)	-3.5 (2.7)
M5d	○	○	○	○	○	Abs. Dif.	-4.2 (2.7)	-4.2 (3.4)	-2.1 (1.5)	-3.2 (1.8)	-4.2 (3.1)	-3.2 (1.8)	-4.2 (3.1)	-3.7 (3.1)
M6	○	○	○	○	○	Raw	-6.2 (4.5)	-6.3 (5.2)	-11.4 (5.1)	-7.2 (4.6)	-8.8 (4.8)	-7.2 (4.6)	-8.8 (4.8)	-8.8 (4.7)
M6d	○	○	○	○	○	Abs. Dif.	-6.7 (5.7)	-6.3 (5.2)	-12.2 (4.8)	-7.8 (5.5)	-9.0 (5.1)	-7.8 (5.5)	-9.0 (5.1)	-8.1 (5.3)
M7	○	○	○	○	○	Raw	-1.1 (1.2)	-1.7 (1.1)	-0.2 (0.7)	-1.1 (1.5)	-0.9 (2.2)	-1.1 (1.5)	-0.9 (2.2)	-1.8 (3.2)
M7d	○	○	○	○	○	Abs. Dif.	-1.3 (1.3)	-1.7 (1.1)	-1.2 (1.1)	-1.0 (1.3)	-0.8 (2.2)	-1.0 (1.3)	-0.8 (2.2)	-1.3 (3.0)
M8	○	○	○	○	○	Raw	0.0 (0.0)	-0.9 (0.7)	-0.8 (0.5)	-1.0 (1.8)	-0.3 (2.5)	-1.0 (1.8)	-0.3 (2.5)	-1.2 (3.2)
M8d	○	○	○	○	○	Abs. Dif.	0.0 (0.4)	0.0 (0.0)	0.0 (0.0)	-1.2 (1.9)	-1.2 (2.8)	-1.2 (1.9)	-1.2 (2.8)	-1.5 (3.5)

Note: The difference in the expected log predictive density (elpd_diff) between the focal model and the model with the smallest LOOIC (i.e. the most appropriate model) is shown. The standard error of elpd_diff (se_diff) is shown in a bracket. If elpd_diff is ≤4, the difference between models is small enough to consider them equivalent. Results are shown according to the difference in the dataset used in the estimation (Dataset 1, 2 or 3) and the number of focal countries in the estimation (i.e. 102 and 127 countries). The multiple logistic regression comprises five explanatory variables, including the species richness of the primary hosts (Dipterocarpaceae) (S1), the species richness of the secondary hosts (*Lithocarpus*) (S2), the geographical distance from Lambir Hill's National Park (GD), the annual mean temperature (AT) and the annual accumulative precipitation (AP). There are two types of explanatory variables, the use of raw value (raw) and the absolute difference between Lambir Hill's National Park and the focal country/region (abs. dif.). The explanatory variables used in each model are shown by open circles.

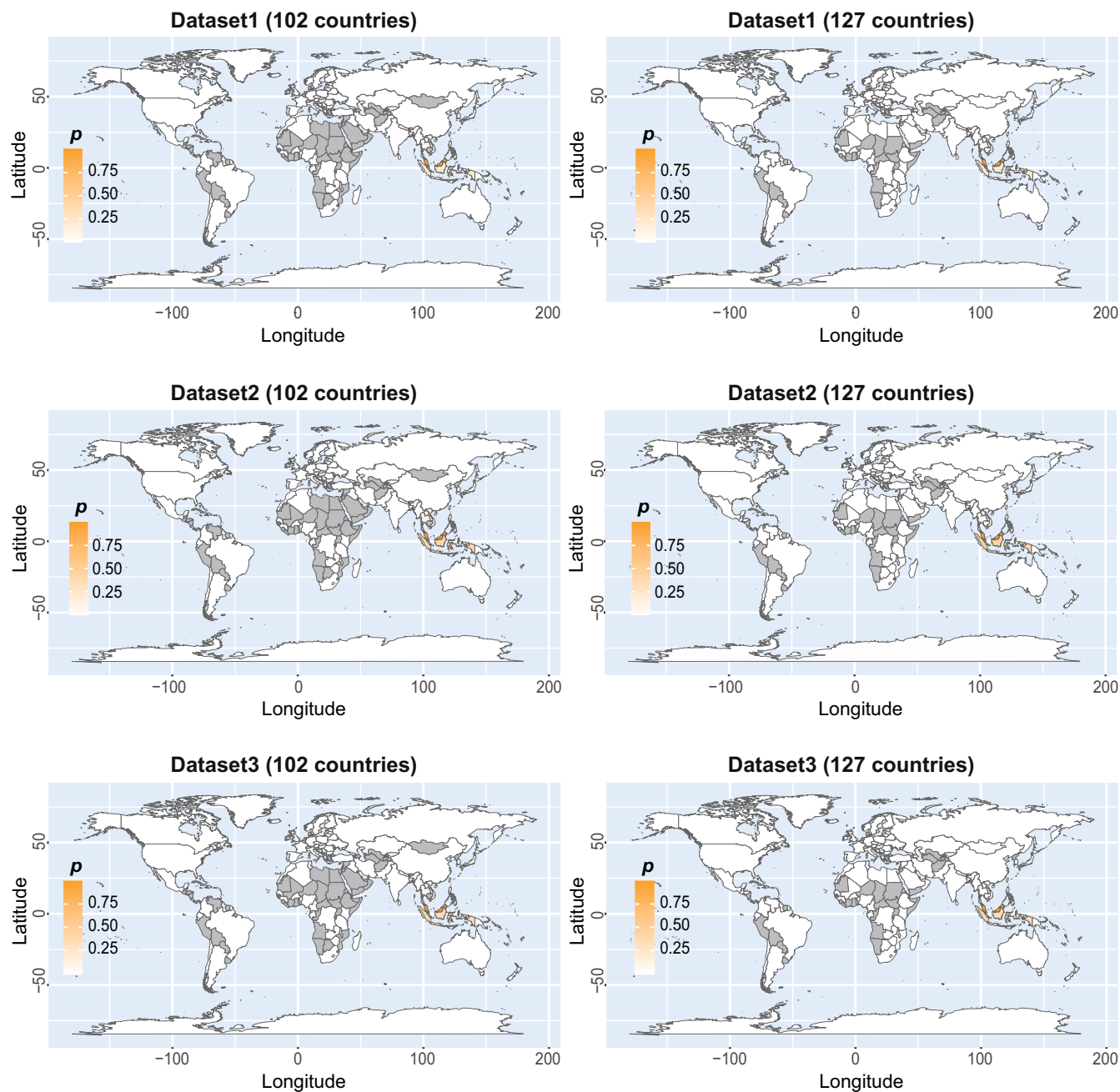


FIGURE 2 Probability of presence in each country of the ectomycorrhizal fungal operational taxonomic units (OTUs) that were detected in Lambir Hill's National Park. Results are shown according to the difference in the dataset used in the estimation (Dataset 1, 2 or 3) and number of focal countries in the estimation (i.e. 102 and 127 countries). The probability of presence in each panel was calculated using the most appropriate model (see Table 2). The probability of presence was estimated at the regional level in the five largest countries in the world (Russia, Canada, China, United States and Brazil). For these countries, the probability is averaged over the regions. The sequence identity threshold between the query OTUs and hit sequences in the database is 97%. The probability of presence is shown by the strength of the orange colour. The grey colour indicates that data are not available.

sufficient availability of preferable hosts is presumably crucial for the expansion of ECM fungi into new areas.

Our findings suggest that dispersal limitation shapes the distinct distance–decay patterns within areas where a sufficient number of preferable hosts are present. In addition to the distribution of the primary host, the geographical distance explains the probability of presence (Figures 1 and S1, Table 2), suggesting that strong

dispersal limitation is present in ECM fungi in dipterocarp rainforests. This pattern is similar to those observed in temperate areas (Glassman et al., 2015; Peay et al., 2012; Peay, Garbelotto, & Bruns, 2010) and is in agreement with the previous finding that distance–decay patterns were especially clear in the tropical ECM fungal communities, in which host densities were low (Bahram et al., 2013). Therefore, it is likely that habitats suitable for ECM fungi are easily fragmented because

of strong host preference, especially in tropical areas, which results in distinct distance–decay patterns (i.e., high endemism and biodiversity) in ECM fungal communities. Overall, the present findings suggest that the presence of preferable hosts determines the potentially habitable areas for ECM fungi, where the dispersal limitation acts as the primary nested factor that controls the distance–decay patterns.

Factors other than preference for Dipterocarpaceae and dispersal limitations seemingly have a limited contribution to the global distribution of ECM fungi in the study site. For instance, both temperature and precipitation were shown to have a limited effect on the probability of presence (Figures 1 and S1, Table 2), which is somewhat inconsistent with previous studies where stronger climatic effects were detected (Delhaye et al., 2024; Li et al., 2023; Talbot et al., 2014; Tedersoo et al., 2014, 2022; van der Linde et al., 2018; Větrovský et al., 2019). This is presumably because the distribution range of Dipterocarpaceae trees is narrower than that of Holarctic ECM trees, such as Pinaceae, Betulaceae, and Fagaceae (Heywood, 1978). In other words, the variation of climatic factors in dipterocarp rainforests is sufficiently small that these factors do not act as the primary determinants for the biogeography of ECM fungi. The results of the modelling also revealed that the distribution of *Lithocarpus* had a limited effect on the probability of occurrence (Figures 1 and S1, Table 2). However, a small fraction of ECM fungi in the study site appeared to be distributed in East Asian countries where Dipterocarpaceae trees are absent but *Lithocarpus* trees are predominant, such as in Japan and Taiwan (Figure 2, Tables 1, S3 and S4), suggesting that the predominance of *Lithocarpus* trees in forests, rather than the species richness, might have a moderate influence on the probability of occurrence. Overall, the biogeography of ECM fungi in dipterocarp rainforests may primarily be explained by preference for Dipterocarpaceae and dispersal limitations, and only partially by preference for *Lithocarpus* (or Fagaceae) and climatic factors.

It remains to be elucidated why host preference is common in ECM fungi. A possible explanation for this is that there might be few disadvantages of host preference at genus or family level in ECM fungi. For instance, most ECM plant genera and families are abundant and widespread (Tedersoo, 2017), and thus the availability of compatible hosts seems to be sufficiently high at local scales even if ECM fungi prefer specific plant genera or families. Moreover, host–fungus compatibility can be often relaxed by the neighbourhood effect, an expansion of ECM colonization from original preferred hosts to neighbouring plants (Bogar & Kennedy, 2013), which may reduce the risk of local extinction of ECM fungal populations (Bogar & Kennedy, 2013; Pérez-Pazos et al., 2021; Tedersoo et al., 2024). Nevertheless, further studies are still needed to verify this hypothesis.

Our results also have important implications for understanding the biodiversity conservation of ECM fungi in dipterocarp rainforests. It is likely that high endemism of ECM fungi is not limited to a dipterocarp rainforest in Lambir Hill's National Park but extends to those in Southeast Asia (Tedersoo et al., 2014, 2022). Thus, ECM fungi in these areas are likely vulnerable to the decline and

fragmentation of dipterocarp rainforests, given that both of host preference and dispersal limitation are the key determinants for the biogeography of ECM fungi. Although further accumulation of cases is needed, our findings suggest high endemism and vulnerability of ECM fungi in tropical rainforests in Borneo.

An important challenge for the future is to separate the effect of host distribution from that of dispersal limitation more clearly. To do this, the focal area of the analysis is needed to be changed to the area where the forest type is different from neighbouring areas, such as tropical montane oak forests in the palaeotropics and plantations of exotic ECM plants. These approaches will be useful to further corroborate the view that the distribution of preferable hosts is the primary determinant for the biogeography of ECM fungi at large scales.

The modelling used in this study is a useful approach to address incomplete and biased sampling for the estimation of the global distribution of ECM fungi; however, some limitations exist. Count matrices of microbial data often contain an excessive number of zero counts, which may be related to undersampling, data redundancy, and a low sequencing depth, resulting in overdispersion (Paulson et al., 2013; Zeng et al., 2022). For instance, about 90% of the variation in ECM fungal communities on a global scale remains unexplained (Tedersoo et al., 2014). Several statistical frameworks have been widely used to address the problem of inflated zeros in regression analysis of count data, such as a zero-inflated Poisson model (Hall, 2000; Lambert, 1992; Zuur et al., 2009), but the problem becomes complicated with the excess zeros in community data matrices. The use of two-step binomial processes, as shown in this study, is an intuitive approach for reducing biases caused by excess zeros observed in global distribution data. We should caution that the registration rates of ITS barcodes are, to a greater or lesser extent, overestimated in countries where only a few database sequences are available (Figure 3, Table S3); however, the accuracy is expected to be improved in future by an increase in ITS barcodes in these countries. A major limitation of this method is the difficulty in including edaphic parameters, which are often important determinants of the ECM fungal biogeography especially at the local scale (Bahram et al., 2015; Mikryukov et al., 2023; Peay, Kennedy, et al., 2010; Talbot et al., 2014; Tedersoo et al., 2014; van der Linde et al., 2018). In addition, we must be cognizant that there are a substantial number of inaccurate sequences (e.g. noisy and chimeric sequences) in the nucleotide sequence database (Peay & Matheny, 2016). Although caution should be exercised for these limitations, the two-step binomial model shown in this study is a powerful tool to address the excess zeros and overdispersion observed in the global distribution data of fungi, which are difficult to overcome in empirical studies of fungal community ecology.

5 | CONCLUSION

Although less attention has been paid to host effects than to dispersal limitations in ECM fungal biogeography, our findings have

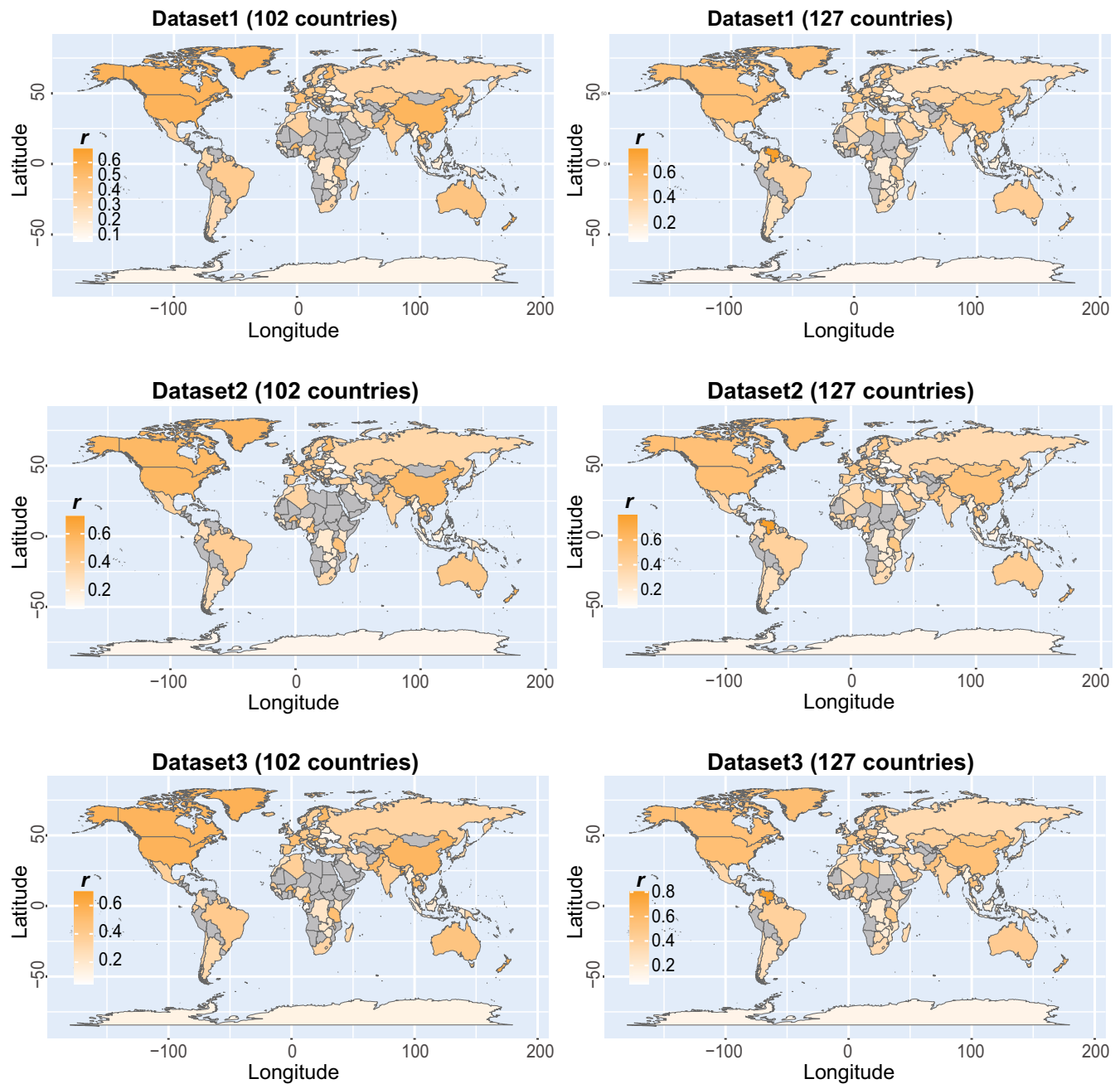


FIGURE 3 Deposition rate of the ectomycorrhizal fungal operational taxonomic units (OTUs) for the nucleotide sequence database for each country. Results are shown according to the difference in the dataset used in the estimation (Dataset 1, 2 and 3) and the number of focal countries in the estimation (i.e. 102 and 127 countries). The deposition rate in each panel was calculated using the most appropriate model (see Table 2). The deposition rate is shown by the strength of the orange colour. The grey colour indicates that data is not available.

important implications for the understanding of how host preference can control the global distribution of ECM fungi. Our results suggest that both host preference and dispersal limitations are the primary determinants of the high endemism of ECM fungi in dipterocarp rainforests. Therefore, it is likely that ECM fungi expand only into areas where preferred hosts are sufficiently available. Within these suitable habitats, dispersal limitations act as a nested factor that controls the distance–decay pattern. Furthermore, the two-step binomial model used in this study is an effective tool to reduce the overdispersion caused by an excessive number of zero counts

and to estimate the fungal distribution patterns at a large scale, although caution should be exercised concerning the limitations, such as the overestimation of ITS registration rates in some countries, the difficulty of using edaphic factors in modelling, and the insufficient reliability of the database sequences.

AUTHOR CONTRIBUTIONS

H.S. designed the research, performed the field surveys, the experiments, developed the modelling, analysed the data, interpreted the results, and wrote the manuscript. L. A. performed the field surveys,

the experiments, analysed the data and interpreted the results. M.T., Y.S., H.B.J., O.B.K. and I. T. performed the field surveys and interpreted the results.

ACKNOWLEDGEMENTS

We thank Dr. Melvin Terry Gumal (Sarawak Forestry Corporation) for providing useful comments on a draft of the manuscript. We are also grateful to Dr. Yoriko Sugiyama (Kyoto University) for supporting the modelling work. This study was conducted in accordance with three Memoranda of Understanding (MOU) signed between the Sarawak Forestry Corporation and the Japan Research Consortium for Tropical Forests in Sarawak (JRCTS) in November 2005 and in February 2021, and signed between the Forest Department Sarawak and JRCTS in November 2012. This work was financially supported by a research grant from a Grant-in-Aid for Scientific Research (20K06796), and JST/JICA, SATREPS (JPMJSA1902).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Input data for the modelling, R commands, and Stan code are available on DataDryad (<https://doi.org/10.5061/dryad.k3j9kd5gs>). A research collaboration was developed with scientists from the countries providing genetic samples, all collaborators are included as co-authors, the results of research have been shared with the provider communities and the broader scientific community (see above), and the research addresses a priority concern, in this case the conservation of organisms being studied. More broadly, our group is committed to international scientific partnerships, as well as institutional capacity building.

ORCID

Hirotooshi Sato  <https://orcid.org/0000-0003-4489-6569>

REFERENCES

- Ashton, P. S. (2005). Lambir's Forest: The world's most dive known tree assemblage? In D. Roubik, S. Sakai, & A. Hamid (Eds.), *Pollination ecology and the rain forest: Sarawak studies* (pp. 191–216). Springer.
- Bahram, M., Koljalg, U., Courty, P. E., Diedhiou, A. G., Kjeller, R., Polme, S., Ryberg, M., Veldre, V., & Tedersoo, L. (2013). The distance decay of similarity in communities of ectomycorrhizal fungi in different ecosystems and scales. *Journal of Ecology*, 101(5), 1335–1344. <https://doi.org/10.1111/1365-2745.12120>
- Bahram, M., Peay, K. G., & Tedersoo, L. (2015). Local-scale biogeography and spatiotemporal variability in communities of mycorrhizal fungi. *New Phytologist*, 205(4), 1454–1463. <https://doi.org/10.1111/nph.13206>
- Balasundaram, S. V., Engh, I. B., Skrede, I., & Kauserud, H. (2015). How many DNA markers are needed to reveal cryptic fungal species? *Fungal Biology*, 119(10), 940–945. <https://doi.org/10.1016/j.funbio.2015.07.006>
- Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K., & Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, 22(3), 148–155. <https://doi.org/10.1016/j.tree.2006.11.004>
- Bogar, L. M., & Kennedy, P. G. (2013). New wrinkles in an old paradigm: Neighborhood effects can modify the structure and specificity of *Alnus*-associated ectomycorrhizal fungal communities. *FEMS Microbiology Ecology*, 83(3), 767–777. <https://doi.org/10.1111/1574-6941.12032>
- Bogar, L. M., & Peay, K. G. (2017). Processes maintaining the coexistence of ectomycorrhizal fungi at a fine spatial scale. In *Biogeography of mycorrhizal symbiosis* (pp. 79–105). Springer.
- Davison, J., Moora, M., Öpik, M., Adholeya, A., Ainsaar, L., Bâ, A., Burla, S., Diedhiou, A. G., Hiiesalu, I., Jairus, T., Johnson, N. C., Kane, A., Koorem, K., Kochar, M., Ndiaye, C., Pärtel, M., Reier, Ü., Saks, Ü., Singh, R., ... Zobel, M. (2015). Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science*, 349(6251), 970–973. <https://doi.org/10.1126/science.aab1161>
- De Wit, R., & Bouvier, T. (2006). 'Everything is everywhere, but, the environment selects'; what did baas Becking and Beijerinck really say? *Environmental Microbiology*, 8(4), 755–758. <https://doi.org/10.1111/j.1462-2920.2006.01017.x>
- Delhay, G., van der Linde, S., Bauman, D., Orme, C. D. L., Suz, L. M., & Bidartondo, M. I. (2024). Ectomycorrhizal fungi are influenced by ecoregion boundaries across Europe. *Global Ecology and Biogeography*, 33(6), e13837. <https://doi.org/10.1111/geb.13837>
- Duarte, S., Bärlocher, F., Pascoal, C., & Cássio, F. (2016). Biogeography of aquatic hyphomycetes: Current knowledge and future perspectives. *Fungal Ecology*, 19, 169–181. <https://doi.org/10.1016/j.funeco.2015.06.002>
- Egidi, E., Delgado-Baquerizo, M., Plett, J. M., Wang, J., Eldridge, D. J., Bardgett, R. D., Maestre, F. T., & Singh, B. K. (2019). A few Ascomycota taxa dominate soil fungal communities worldwide. *Nature Communications*, 10(1), 2369. <https://doi.org/10.1038/s41467-019-10373-z>
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472. <https://doi.org/10.1214/ss/1177011136>
- Glassman, S. I., Peay, K. G., Talbot, J. M., Smith, D. P., Chung, J. A., Taylor, J. W., Vilgalys, R., & Bruns, T. D. (2015). A continental view of pine-associated ectomycorrhizal fungal spore banks: A quiescent functional guild with a strong biogeographic pattern. *New Phytologist*, 205(4), 1619–1631. <https://doi.org/10.1111/nph.13240>
- Glynou, K., Ali, T., Buch, A. K., Haghi Kia, S., Ploch, S., Xia, X., Çelik, A., Thines, M., & Maciá-Vicente, J. G. (2016). The local environment determines the assembly of root endophytic fungi at a continental scale. *Environmental Microbiology*, 18(8), 2418–2434. <https://doi.org/10.1111/1462-2920.13112>
- Gupta, A. K., & Nadarajah, S. (2004). *Handbook of beta distribution and its applications*. CRC Press.
- Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics*, 56(4), 1030–1039. <https://doi.org/10.1111/j.0006-341x.2000.01030.x>
- Heywood, V. H. (1978). *Flowering plants of the world*. Oxford University Press.
- Hijmans, R. J., Barbosa, M., Ghosh, A., & Mandel, A. (2023). *Geodata: Download geographic data. R package version 0.5*.
- Hijmans, R. J., Van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., Lamigueiro, O. P., Bevan, A., Racine, E. B., & Shortridge, A. (2015). *Package 'raster'*.
- Holt, B. G., Lessard, J.-P., Borregaard, M. K., Fritz, S. A., Araújo, M. B., Dimitrov, D., Fabre, P.-H., Graham, C. H., Graves, G. R., Jönsson, K. A., Nogués-Bravo, D., Wang, Z., Whittaker, R. J., Fjeldså, J., Rahbek, C., & Jönsson, K. A. (2013). An update of Wallace's zoogeographic regions of the world. *Science*, 339(6115), 74–78. <https://doi.org/10.1126/science.1228282>
- Ishida, T. A., Nara, K., & Hogetsu, T. (2007). Host effects on ectomycorrhizal fungal communities: Insight from eight host species in

- mixed conifer-broadleaf forests. *New Phytologist*, 174(2), 430–440. <https://doi.org/10.1111/j.1469-8137.2007.02016.x>
- Kennedy, P. (2010). Ectomycorrhizal fungi and interspecific competition: Species interactions, community structure, coexistence mechanisms, and future research directions. *New Phytologist*, 187(4), 895–910. <https://doi.org/10.1111/j.1469-8137.2010.03399.x>
- Kennedy, P. G., Peay, K. G., & Bruns, T. D. (2009). Root tip competition among ectomycorrhizal fungi: Are priority effects a rule or an exception? *Ecology*, 90(8), 2098–2107. <https://doi.org/10.1890/08-1291.1>
- Kohout, P. (2017). Biogeography of ericoid mycorrhiza. In L. Tedersoo (Ed.), *Biogeography of mycorrhizal symbiosis* (pp. 179–193). Springer International Publishing.
- Köljal, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F., Bahram, M., Bates, S. T., Bruns, T. D., Bengtsson-Palme, J., Callaghan, T. M., Douglas, B., Drenkhan, T., Eberhardt, U., Dueñas, M., Grebenc, T., Griffith, G. W., Hartmann, M., Kirk, P. M., Kohout, P., ... Callaghan, T. M. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology*, 22(21), 5271–5277. <https://doi.org/10.1111/mec.12481>
- Kreft, H., & Jetz, W. (2010). A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography*, 37(11), 2029–2053. <https://doi.org/10.1111/j.1365-2699.2010.02375.x>
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34, 1–14. <https://doi.org/10.2307/1269547>
- Li, P., Tedersoo, L., Crowther, T. W., Wang, B., Shi, Y., Kuang, L., Li, T., Wu, M., Liu, M., Luan, L., Liu, J., Li, D., Li, Y., Wang, S., Saleem, M., Dumbrell, A. J., Li, Z., & Luan, L. (2023). Global diversity and biogeography of potential phytopathogenic fungi in a changing world. *Nature Communications*, 14(1), 6482. <https://doi.org/10.1038/s41467-023-42142-4>
- Mikryukov, V., Dulya, O., Zizka, A., Bahram, M., Hagh-Doust, N., Anslan, S., Prylutskiy, O., Delgado-Baquerizo, M., Maestre, F. T., Nilsson, H., Pärn, J., Öpik, M., Moora, M., Zobel, M., Espenberg, M., Mander, Ü., Khalid, A. N., Corrales, A., Agan, A., ... Nilsson, H. (2023). Connecting the multiple dimensions of global soil fungal diversity. *Science Advances*, 9(48), ead8016. <https://doi.org/10.1126/sciadv.adj8016>
- Molina, R., & Horton, T. R. (2015). Mycorrhiza specificity: Its role in the development and function of common mycelial networks. In T. R. Horton (Ed.), *Mycorrhizal networks* (pp. 1–39). Springer.
- Molina, R., Massicotte, H., & Trappe, J. M. (1992). Specificity phenomena in mycorrhizal symbioses: Community ecological consequences and practical applications. In A. Allen (Ed.), *Mycorrhizal functioning* (pp. 357–422). Chapman and Hall.
- Morrone, J. J. (2018). The spectre of biogeographical regionalization. *Journal of Biogeography*, 45(2), 282–288. <https://doi.org/10.1111/jbi.13135>
- Nilsson, R. H., Larsson, K.-H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F. O., Tedersoo, L., Saar, I., Köljal, U., & Tedersoo, L. (2018). The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research*, 47, D259–D264. <https://doi.org/10.1093/nar/gky1022>
- Paulson, J. N., Stine, O. C., Bravo, H. C., & Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. *Nature Methods*, 10(12), 1200–1202. <https://doi.org/10.1038/nmeth.2658>
- Peay, K. G., Garbelotto, M., & Bruns, T. D. (2010). Evidence of dispersal limitation in soil microorganisms: Isolation reduces species richness on mycorrhizal tree islands. *Ecology*, 91(12), 3631–3640. <https://doi.org/10.1890/09-2237.1>
- Peay, K. G., Kennedy, P. G., Davies, S. J., Tan, S., & Bruns, T. D. (2010). Potential link between plant and fungal distributions in a dipterocarp rainforest: Community and phylogenetic structure of tropical ectomycorrhizal fungi across a plant and soil ecotone. *New Phytologist*, 185(2), 529–542. <https://doi.org/10.1111/j.1469-8137.2009.03075.x>
- Peay, K. G., & Matheny, P. B. (2016). The biogeography of ectomycorrhizal fungi—a history of life in the subterranean. In F. Martin (Ed.), *Molecular mycorrhizal symbiosis* (pp. 341–361). Wiley-Blackwell.
- Peay, K. G., Russo, S. E., McGuire, K. L., Lim, Z., Chan, J. P., Tan, S., & Davies, S. J. (2015). Lack of host specificity leads to independent assortment of dipterocarps and ectomycorrhizal fungi across a soil fertility gradient. *Ecology Letters*, 18(8), 807–816. <https://doi.org/10.1111/ele.12459>
- Peay, K. G., Schubert, M. G., Nguyen, N. H., & Bruns, T. D. (2012). Measuring ectomycorrhizal fungal dispersal: Macroecological patterns driven by microscopic propagules. *Molecular Ecology*, 21(16), 4122–4136. <https://doi.org/10.1111/j.1365-294x.2012.05666.x>
- Pérez-Pazos, E., Certano, A., Gagne, J., Lebeuf, R., Siegel, N., Nguyen, N., & Kennedy, P. G. (2021). The slippery nature of ectomycorrhizal host specificity: Suillus fungi associated with novel pinoid (Picea) and abietoid (Abies) hosts. *Mycologia*, 113(5), 891–901. <https://doi.org/10.1080/00275514.2021.1921525>
- Pölme, S., Bahram, M., Yamanaka, T., Nara, K., Dai, Y. C., Grebenc, T., Kraigher, H., Toivonen, M., Wang, P.-H., Matsuda, Y., Naadel, T., Kennedy, P. G., Köljal, U., & Tedersoo, L. (2013). Biogeography of ectomycorrhizal fungi associated with alders (Alnus spp.) in relation to biotic and abiotic variables at the global scale. *New Phytologist*, 198(4), 1239–1249. <https://doi.org/10.1111/nph.12170>
- Queloz, V., Sieber, T. N., Holdenrieder, O., McDonald, B. A., & Grünig, C. R. (2011). No biogeographical pattern for a root associated fungal species complex. *Global Ecology and Biogeography*, 20, 160–169. <https://doi.org/10.1111/j.1466-8238.2010.00589.x>
- R Core Team. (2023). *R: A language and environment for statistical Computing*. R Foundation for Statistical Computing.
- Rinaldi, A. C., Comandini, O., & Kuyper, T. W. (2008). Ectomycorrhizal fungal diversity: Separating the wheat from the chaff. *Fungal Diversity*, 33, 1–45.
- Sato, H. (2024). The evolution of ectomycorrhizal symbiosis in the late cretaceous is a key driver of explosive diversification in agaricomycetes. *New Phytologist*, 241(1), 444–460. <https://doi.org/10.1111/nph.19055>
- Sato, H., Tanabe, A. S., & Toju, H. (2015). Contrasting diversity and host association of ectomycorrhizal Basidiomycetes versus root-associated ascomycetes in a dipterocarp rainforest. *PLoS One*, 10(4), e0125550. <https://doi.org/10.1371/journal.pone.0125550>
- Sato, H., Tsujino, R., Kurita, K., Yokoyama, K., & Agata, K. (2012). Modelling the global distribution of fungal species: New insights into microbial cosmopolitanism. *Molecular Ecology*, 21, 5599–5612. <https://doi.org/10.1111/mec.12053>
- Schmidt, P.-A., Bálint, M., Greshake, B., Bandow, C., Römbke, J., & Schmitt, I. (2013). Illumina metabarcoding of a soil fungal community. *Soil Biology and Biochemistry*, 65, 128–132. <https://doi.org/10.1016/j.soilbio.2013.05.014>
- Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., Chen, W., & Fungal Barcoding Consortium; Fungal Barcoding Consortium Author List. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proceedings of the National Academy of Sciences*, 109(16), 6241–6246. <https://doi.org/10.1073/pnas.1117018109>
- Smith, M. E., Douhan, G. W., & Rizzo, D. M. (2007). Intra-specific and intra-sporocarp ITS variation of ectomycorrhizal fungi as assessed by rDNA sequencing of sporocarps and pooled ectomycorrhizal roots from a Quercus woodland. *Mycorrhiza*, 18(1), 15–22. <https://doi.org/10.1007/s00572-007-0148-z>
- Smith, S. E., & Read, D. J. (2008). *Mycorrhizal symbiosis, third Edition*. Academic Press.
- Stan-Development-Team. (2023). *RStan: The R interface to Stan. R package version 2.26.23*.

- Stürmer, S. L., Bever, J. D., & Morton, J. B. (2018). Biogeography of arbuscular mycorrhizal fungi (Glomeromycota): A phylogenetic perspective on species distribution patterns. *Mycorrhiza*, 28, 587–603. <https://doi.org/10.1007/s00572-018-0864-6>
- Talbot, J. M., Bruns, T. D., Taylor, J. W., Smith, D. P., Branco, S., Glassman, S. I., Erlandson, S., Vilgalys, R., Liao, H.-L., Smith, M. E., & Peay, K. G. (2014). Endemism and functional convergence across the north American soil mycobiome. *Proceedings of the National Academy of Sciences*, 111(17), 6341–6346. <https://doi.org/10.1073/pnas.1402584111>
- Tanabe, A. S., & Toju, H. (2013). Two new computational methods for universal DNA barcoding: A benchmark using barcode sequences of bacteria, archaea, animals, fungi, and land plants. *PLoS One*, 8(10), e76910. <https://doi.org/10.1371/journal.pone.0152242>
- Tedersoo, L. (2017). *Biogeography of mycorrhizal symbiosis*. Springer International Publishing.
- Tedersoo, L., Bahram, M., Pölme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., Ruiz, L. V., Vasco-Palacios, A. M., Thu, P. Q., Suija, A., Smith, M. E., Sharp, C., Saluveer, E., Saitta, A., Rosas, M., Riit, T., Ratkowsky, D., Pritsch, K., Põldmaa, K., ... Suija, A. (2014). Global diversity and geography of soil fungi. *Science*, 346(6213), 1256688. <https://doi.org/10.1126/science.1256688>
- Tedersoo, L., Drenkhan, R., Abarenkov, K., Anslan, S., Bahram, M., Biteniek, K., Buegger, F., Gohar, D., Hagh-Doust, N., Klavina, D., Makovskis, K., Zusevica, A., Pritsch, K., Padari, A., Pölme, S., Rahimlou, S., Rungis, D., & Klavina, D. (2024). The influence of tree genus, phylogeny, and richness on the specificity, rarity, and diversity of ectomycorrhizal fungi. *Environmental Microbiology Reports*, 16(2), e13253. <https://doi.org/10.1111/1758-2229.13253>
- Tedersoo, L., Jairus, T., Horton, B. M., 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(2), 479–490. <https://doi.org/10.1111/j.1469-8137.2008.02561.x>
- Tedersoo, L., May, T. W., & Smith, M. E. (2010). Ectomycorrhizal lifestyle in fungi: Global diversity, distribution, and evolution of phylogenetic lineages. *Mycorrhiza*, 20(4), 217–263. <https://doi.org/10.1007/s00572-009-0274-x>
- Tedersoo, L., Mikryukov, V., Zizka, A., Bahram, M., Hagh-Doust, N., Anslan, S., Prylutsky, O., Delgado-Baquerizo, M., Maestre, F. T., Pärn, J., Öpik, M., Moora, M., Zobel, M., Espenberg, M., Mander, Ü., Khalid, A. N., Corrales, A., Agan, A., Vasco-Palacios, A. M., ... Pärn, J. (2022). Global patterns in endemism and vulnerability of soil fungi. *Global Change Biology*, 28(22), 6696–6710. <https://doi.org/10.1111/gcb.16398>
- Tedersoo, L., & Nara, K. (2010). General latitudinal gradient of biodiversity is reversed in ectomycorrhizal fungi. *New Phytologist*, 185(2), 351–354. <https://doi.org/10.1111/j.1469-8137.2009.03134.x>
- Tedersoo, L., Sadam, A., Zambrano, M., Valencia, R., & Bahram, M. (2009). Low diversity and high host preference of ectomycorrhizal fungi in Western Amazonia, a neotropical biodiversity hotspot. *The ISME Journal*, 4(4), 465–471. <https://doi.org/10.1038/ismej.2009.131>
- Tedersoo, L., & Smith, M. E. (2013). Lineages of ectomycorrhizal fungi revisited: Foraging strategies and novel lineages revealed by sequences from belowground. *Fungal Biology Reviews*, 27(3), 83–99. <https://doi.org/10.1016/j.fbr.2013.09.001>
- van der Linde, S., Suz, L. M., Orme, C. D. L., Cox, F., Andreae, H., Asi, E., Atkinson, B., Benham, S., Carroll, C., Cools, N., De Vos, B., Dietrich, H. P., Eichhorn, J., Gehrmann, J., Grebenc, T., Gweon, H. S., Hansen, K., Jacob, F., Kristöfel, F., ... Cools, N. (2018). Environment and host as large-scale controls of ectomycorrhizal fungi. *Nature*, 558(7709), 243–248. <https://doi.org/10.1038/s41586-018-0189-9>
- Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27, 1413–1432. <https://doi.org/10.1007/s11222-016-9696-4>
- Vehtari, A., Gelman, A., Gabry, J., & Yao, Y. (2021). Package 'loo'. *Efficient leave-one-out cross-validation and WAIC for Bayesian models*. <https://discourse.mc-stan.org/>
- Větrovský, T., Kohout, P., Kopecký, M., Machac, A., Man, M., Bahnmann, B. D., Brabcová, V., Choi, J., Meszárošová, L., Human, Z. R., Lepinay, C., Lladó, S., López-Mondéjar, R., Martinović, T., Mašínová, T., Morais, D., Navrátilová, D., Od, I., ... Human, Z. R. (2019). A meta-analysis of global fungal distribution reveals climate-driven patterns. *Nature Communications*, 10(1), 5142. <https://doi.org/10.1038/s41467-019-13164-8>
- Větrovský, T., Morais, D., Kohout, P., Lepinay, C., Algora, C., Hollá, S. A., Bahnmann, B. D., Bílohněd, K., Brabcová, V., D'Alò, F., Human, Z. R., Jomura, M., Kolařík, M., Kvasničková, J., Lladó, S., López-Mondéjar, R., Martinović, T., Mašínová, T., ... D'Alò, F. (2020). GlobalFungi, a global database of fungal occurrences from high-throughput-sequencing metabarcoding studies. *Scientific Data*, 7(1), 1–14. <https://doi.org/10.1038/s41597-020-0567-7>
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(12), 3571–3594. <https://doi.org/10.48550/arXiv.1004.2316>
- Werth, S. (2011). Biogeography and phylogeography of lichen fungi and their photobionts. In D. Fontaneto (Ed.), *Biogeography of microscopic organisms: Is everything small everywhere* (pp. 191–208). Cambridge University Press.
- Zeng, Y., Li, J., Wei, C., Zhao, H., & Wang, T. (2022). mbDenoise: Microbiome data denoising using zero-inflated probabilistic principal components analysis. *Genome Biology*, 23(1), 94. <https://doi.org/10.1186/s13059-022-02657-3>
- Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R* (Vol. 574). Springer.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Sato, H., Lain, A., Mizuno, T., Yamashita, S., Hassan, J. B., Othman, K. B., & Itioka, T. (2024). Host preference explains the high endemism of ectomycorrhizal fungi in a dipterocarp rainforest. *Molecular Ecology*, 00, e17529. <https://doi.org/10.1111/mec.17529>