

# Tropical tree ectomycorrhiza are distributed independently of soil nutrients

Received: 17 March 2023

Accepted: 1 December 2023

Published online: 10 January 2024

 Check for updates

A list of authors and their affiliations appears at the end of the paper

Mycorrhizae, a form of plant–fungal symbioses, mediate vegetation impacts on ecosystem functioning. Climatic effects on decomposition and soil quality are suggested to drive mycorrhizal distributions, with arbuscular mycorrhizal plants prevailing in low-latitude/high-soil-quality areas and ectomycorrhizal (EcM) plants in high-latitude/low-soil-quality areas. However, these generalizations, based on coarse-resolution data, obscure finer-scale variations and result in high uncertainties in the predicted distributions of mycorrhizal types and their drivers. Using data from 31 lowland tropical forests, both at a coarse scale (mean-plot-level data) and fine scale (20 × 20 metres from a subset of 16 sites), we demonstrate that the distribution and abundance of EcM-associated trees are independent of soil quality. Resource exchange differences among mycorrhizal partners, stemming from diverse evolutionary origins of mycorrhizal fungi, may decouple soil fertility from the advantage provided by mycorrhizal associations. Additionally, distinct historical biogeographies and diversification patterns have led to differences in forest composition and nutrient-acquisition strategies across three major tropical regions. Notably, Africa and Asia's lowland tropical forests have abundant EcM trees, whereas they are relatively scarce in lowland neotropical forests. A greater understanding of the functional biology of mycorrhizal symbiosis is required, especially in the lowland tropics, to overcome biases from assuming similarity to temperate and boreal regions.

Many plants establish symbiotic relationships with soil microbes, enabling them to access soil resources that would otherwise be unavailable or to gain protection against biotic and abiotic stress<sup>1</sup>. One largely recognized form of symbiosis is the association between the majority of vascular plants and mycorrhizal fungi, which occurs in or on the roots and is known as mycorrhiza<sup>2</sup>. Mycorrhizal fungi can improve plant mineral nutrition<sup>3</sup>, stress tolerance (for example, to drought) and defence (for example, to soil-borne pathogens)<sup>4</sup>. In exchange, the host plant provides the fungus with the carbon required for functioning<sup>5</sup>. Most trees associate with either arbuscular mycorrhizal (AM; but refer to refs. 6,7) or ectomycorrhizal (EcM) fungi, forming the AM and EcM types, respectively. AM predominate in the number of host tree species—about 72% of vascular plant species are AM and about 2% are

EcM<sup>8</sup>. AM and EcM types differ in their nutrient economies and effects on soil biogeochemistry<sup>3,9</sup>, and variations in their relative dominance may have large-scale consequences for ecosystem functioning and biogeochemical cycles<sup>10,11</sup>. Therefore, accurate characterization of the distribution of mycorrhizal types and the factors that drive their distribution is critical to understanding and modelling forest biogeochemistry and, ultimately, climate feedbacks.

Differences in traits between AM and EcM fungi and between AM and EcM-associating trees suggest differences in their nutrient economies and foraging strategies. It is generally accepted that EcM fungi have a higher capacity to mobilize nitrogen (N) and phosphorus (P) from soil organic matter than AM fungi<sup>3,9</sup>. The high capacity of EcM fungi to exploit soil organic matter can lead to a reduction in

✉ e-mail: [jamedinavega@gmail.com](mailto:jamedinavega@gmail.com)

litter decomposition (the Gadgil effect<sup>12</sup>), as EcM fungi obtain carbon from their host plant and effectively compete with decomposers such as saprotrophic fungi, which rely on decomposing organic matter for carbon. Some AM fungi can also acquire N<sup>13</sup> and P<sup>14</sup> from organic sources by recruiting and hosting distinct bacteria in their extraradical hyphae (the hyphosphere microbiome). Whereas EcM fungi are mostly recognized for their high capacity to mobilize N and P in nutrient-limited and organic soils<sup>3,15</sup>, they, like AM fungi, also contribute to the mobilization of other nutrients such as potassium (K), magnesium (Mg), calcium (Ca), sulfur (S), iron (Fe), zinc (Zn), copper (Cu) and manganese (Mn)<sup>16</sup>. EcM trees may have longer-lived leaves and more recalcitrant litter, leading to a slower nutrient economy than AM trees<sup>17</sup>. The combination of the poor-quality litter of EcM trees and the high capability of EcM fungi to acquire nutrients from it may result in a positive feedback between slow decomposition rates and increased soil organic matter accumulation<sup>9</sup>. In such forests, it is predicted that a high accumulation of organic matter and a low availability of nutrients in inorganic form may ultimately favour and lead to a dominance of EcM (the mycorrhizal-associated nutrient economy framework in ref. 9). Differences in the nutrient economies and foraging strategies between EcM and AM fungi and between AM and EcM-associating trees may lead to divergent costs and benefits along spatial environmental gradients (but refer to ref. 18), which may result in different distributions.

The dominant hypothesis for global patterns in mycorrhizal symbiosis is that the relative abundance of EcM decreases along a gradient from cold and/or dry to warm, wet climates through the effects of climate on soil development and decomposition rates<sup>19,20</sup>. A recent quantitative approximation of the global distribution of forest–tree mycorrhizal symbionts found that EcM dominate at higher latitudes (and higher elevations), where generally lower temperatures and precipitation lead to slower decomposition rates and lower availability of nutrients in inorganic forms<sup>19</sup>. Conversely, AM were found to dominate at lower latitudes (and lower elevations), where warmer, wetter environments lead to faster decomposition rates and higher availability of nutrients in inorganic forms<sup>19</sup>. These global-scale approximations of the distribution of AM and EcM appear to be consistent with their different nutrient economies<sup>3,9</sup>. However, these approximations are based on a coarse resolution of global vegetation, climate and soil patterns and obscure large variations at finer scales and dramatically underrepresent tropical forests. As a result, there remain high uncertainties in both the predicted distributions of mycorrhizal types<sup>21,22</sup> and the drivers of those distributions.

In tropical forests, the dominance of EcM trees generally increases with elevation<sup>19</sup>. In lowland tropical forests, EcM associations are often described as patchy and rare<sup>8</sup> and are generally suggested to be found in forests characterized by single-species dominance, deep leaf litter and nutrient-poor sites<sup>15,19,23</sup>. However, known counter-examples to this generalization include tropical lowland forests dominated by many EcM tree species (for example, the diverse mixed dipterocarp forests in Southeast Asia<sup>24</sup> or mixed-legume forests in tropical Africa<sup>25</sup>), a lack of association between the distribution of EcM host trees and soil chemical properties<sup>23</sup> and forests dominated by single AM species (for example, *Mora excelsa* Benth. (Fabaceae))<sup>26</sup>. Additionally, the differences in traits between EcM and AM tree species and between EcM and AM fungal taxa supporting contrasting mycorrhizal nutrient economies or foraging strategies do not always hold. EcM fungi do not always improve plant nutrient uptake in comparison to AM fungi when grown in the same soil medium<sup>27</sup>, and foliar traits of AM and EcM trees do not always differ in the expected ways<sup>28</sup>. These findings challenge the existence of a universally applicable mycorrhizal-associated nutrient economy, particularly in lowland tropical forests (refs. 29,30 for examples in South American temperate regions), and suggest that the drivers that shape their distribution need to be re-examined.

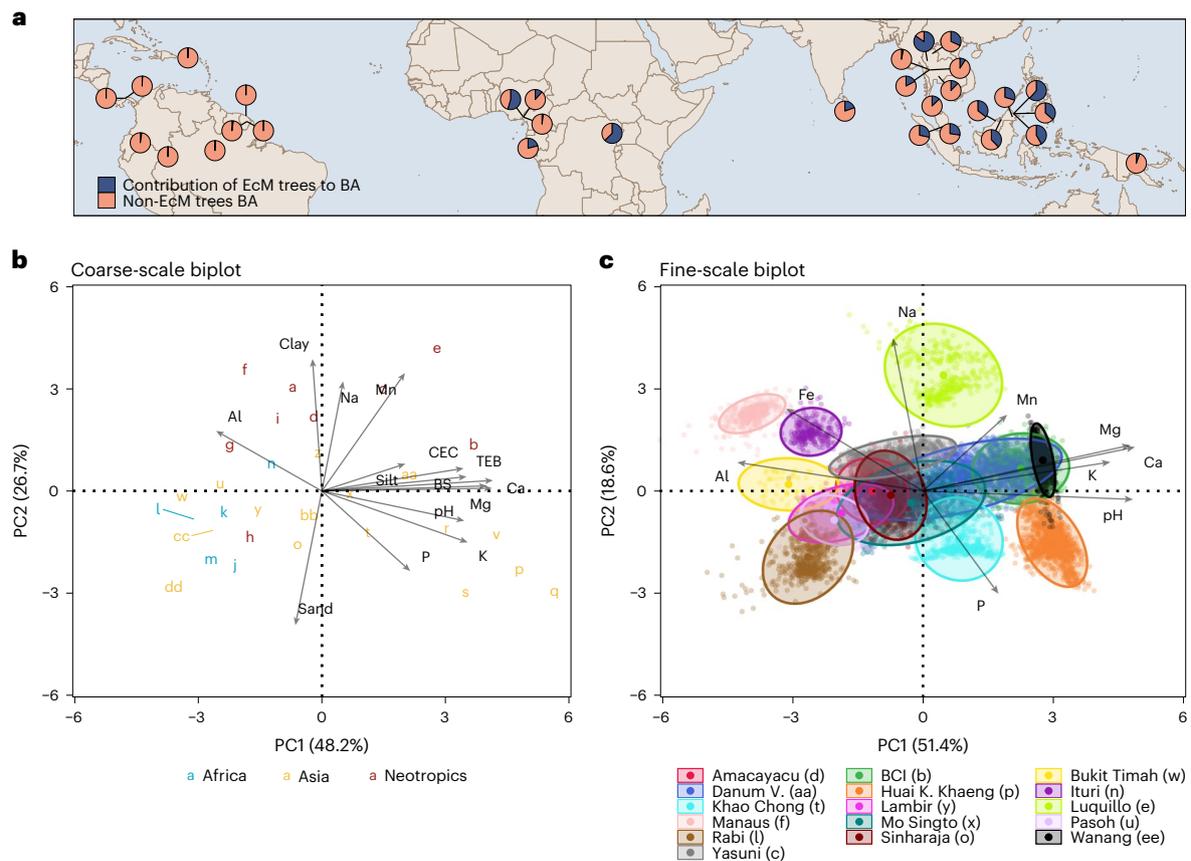
We evaluated the applicability of the mycorrhizal-associated nutrient economy framework across lowland terra firme tropical

forests by testing the hypothesis that the distribution and abundance of EcM trees in lowland tropical forests are related to variations in soil properties. We focused on EcM trees because the abundance of EcM tree and AM tree individuals are generally inversely related<sup>19,31</sup>. We compiled data on the relative abundance of EcM trees (the proportion of basal area (BA) contributed by EcM trees) and soil properties of 31 lowland tropical forests from the Forest Global Earth Observatory (ForestGEO) plot network of research sites<sup>32</sup> and the literature (Fig. 1a, Extended Data Table 1 and Extended Data Fig. 1). We created two datasets. The first dataset contained mean-plot-level data on the relative abundance of EcM trees, soil chemistry (Al, Ca, K, Mg, Mn, Na, CEC (cation exchange capacity) and TEB (total exchangeable bases) in  $\text{cmol}_c \text{kg}^{-1}$ , plant-available P in  $\text{mg kg}^{-1}$ , pH and BS (percent base saturation)) and soil texture (the proportion of sand, clay and silt content) for 30 sites from three regions: neotropics, Africa and Asia. A single site from Oceania was excluded from this mean-plot-level dataset due to sample size (Methods; statistical analyses for the coarse-scale data). The second dataset contained the fine-scale data (quadrat-level;  $20 \times 20 \text{ m}$ ) on the relative abundance of EcM trees and soil chemical properties for 16 sites (Extended Data Table 2), which were the maximum number of sites with all trees  $\geq 1 \text{ cm}$  in diameter at breast height (DBH) identified and with the most complete and consistently measured set of soil variables, including Al, Ca, K, Mg, Na, Fe, Mn, plant-available P and pH. Refer to methods for a full description of the datasets.

We used principal component analysis (PCA) to construct gradients in soil properties at both coarse (using mean-plot-level soil data) and fine (using soil data at every  $20 \times 20 \text{ m}$  quadrat) scales. Both PCAs revealed similar patterns (Fig. 1b,c). The first principal component (PC) of both PCAs described variation in soil chemical properties (soil nutrient availability) and increased with increasing concentrations of soil bases and pH and with decreasing Al (Extended Data Table 3). The second PC was comparable at both fine and coarse scales. The second PC described variation in soil physical properties and P availability and increased with clay content (Fig. 1b) and Na (Fig. 1b,c) concentration and with decreasing sand content (Fig. 1b) and P concentration (Fig. 1b,c). We included the two PCs from both PCAs as covariates in both a generalized linear model (GLM) and a generalized linear mixed effects model (GLMM) to assess the association between EcM tree abundance in BA and soil properties. The GLM was used to analyse coarse-scale associations across forests, whereas the GLMM was used to analyse fine-scale associations within and among forests. Given a high presence of zeroes in the fine-scale dataset (Extended Data Table 2), the GLMM was a joint model with two components, a discrete component to assess if an event occurs (the probability of observing EcM trees) and a continuous component to assess the event's intensity given that it occurs (the relative abundance in BA of EcM trees conditional on their occurrence; Methods)<sup>33,34</sup>. We included quadrat-level topography (elevation, slope and convexity) and (total) BA in the analysis on fine-scale data to characterize the terrain and to control for differences in exposure (a higher opportunity of observing EcM trees and higher relative abundance of EcM trees in quadrats with larger BAs or the opposite), respectively.

## Results and discussion

The relative abundance in BA of EcM trees in lowland tropical forests exhibited high variability and was unrelated to the variation in soil properties at both coarse and fine scales (Figs. 2 and 3). At the coarse scale (using mean-plot level data), the relative abundance of EcM trees ranged from 0.02% to 84.5% BA (mean 22%), being lower in the neotropics (range = 0.02–2.3%, mean = 0.8) than the Afrotropics (range = 2.3–62.6%, mean = 30.4) or Southeast Asia (range = 3.2–84.5%, mean = 30.5). Additionally, the relative abundance of EcM trees at the coarse scale was independent of soil variation among and within these three major tropical regions (Fig. 2 and Extended Data Table 4).



**Fig. 1 | Study sites and variation in the relative abundance of EcM trees and soil properties.** **a**, Location of the study sites and the relative contribution of EcM trees to BA. Plot identities are mapped in Extended Data Fig. 1. **b, c**, Biplots from two PCAs. PC1 and PC2 indicate the first and second PCs for each PCA, respectively. Biplot in **b** uses mean soil plot-level data (coarse scale) whereas the biplot in **c** uses spatially detailed (quadrat-level;  $20 \times 20$  m) soil data (fine scale)

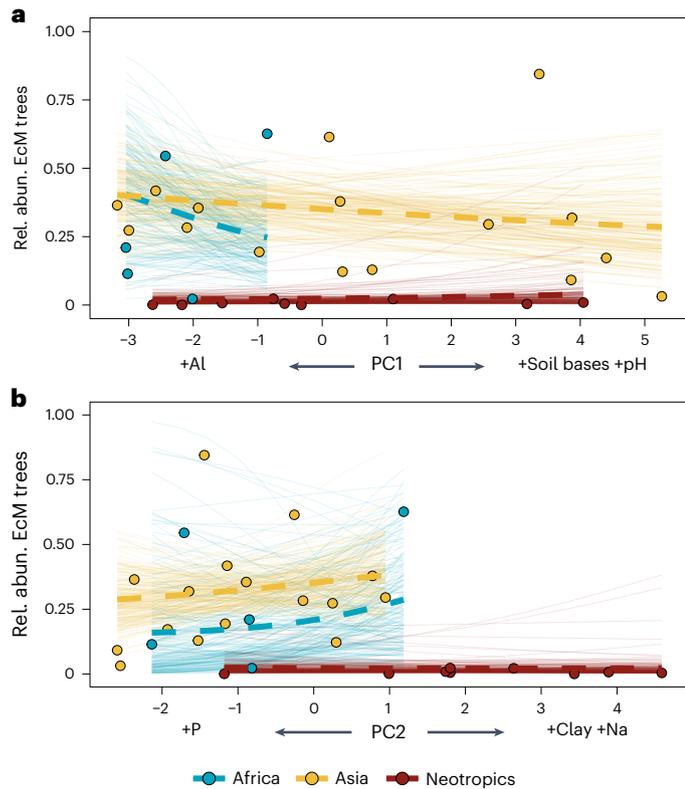
from 30 and 16 lowland tropical forests, respectively. In **b**, letters indicate forest identity (Extended Data Fig. 1) and are colour-coded by major tropical region (neotropics, tropical Africa and tropical Asia). In **c**, each dot indicates a  $20 \times 20$  m quadrat and is colour-coded by forest. Ellipses in **c** concentrate 95% of quadrats from each forest. Vectors in **b** and **c** illustrate the soil variables included in each PCA. Map made with Natural Earth.

At the fine scale ( $20 \times 20$  m quadrats), the probability of observing EcM trees and the relative abundance of those EcM trees were also independent of both PC1 and PC2 (Fig. 3 and Extended Data Table 5). Although there was a decline in the probability of observing EcM trees with an increase in the availability of nutrients in inorganic form (PC1, Fig. 3a), the decline was small. That is, even in quadrats with higher availability of nutrients, the mean probability of observing an EcM tree remained above 0.75. The observed decline in the probability of observing EcM trees with increasing availability of soil nutrients was primarily driven by the low presence of EcM trees in high-fertility quadrats within three out of 16 sites (Fig. 3c). The three sites, namely, Danum Valley in Malaysia, Khao Chong in Thailand and Amacayacu in Colombia, exhibit 2%, 10% and 48% of quadrats without EcM trees, respectively (Supplementary Discussion). After excluding these three sites from the analysis to evaluate their impact on the observed negative association between the probability of observing EcM trees and PC1 at the fine scale, the negative association disappeared (Extended Data Table 6). Moreover, when the analysis at a fine scale was performed with  $40 \times 40$  m quadrats (Methods), the negative association between the probability of observing EcM trees and PC1 was no longer evident, indicating the robustness of our analysis across spatial scales.

Our results are consistent with studies from specific lowland tropical forests (but refer to ref. 35). For example, in the southern part of Korup National Park, Cameroon, areas with a higher abundance of EcM trees are not associated with lower concentrations of soil nutrients (P and N)<sup>36</sup> (data from this study site was included in our analysis

at a coarse scale; Extended Data Table 1). In Guyana, transect surveys indicate that forests with high dominance of the leguminous EcM tree species *Dicymbe corymbosa* and *D. altsonii* have high variability in soil texture and macronutrients<sup>37</sup>, suggesting a lack of association between soil attributes with the local distribution and abundance of EcM trees (also ref. 23). On the contrary, the EcM BA in a forest in Malaysia was higher on less fertile soil types, which may be due to the overall higher BA on those areas<sup>31</sup>, and is consistent with our finding that the proportion of EcM BA increases with the total BA (Extended Data Table 5), particularly in SE Asia (Extended Data Fig. 2).

Our analysis demonstrates that EcM trees in lowland tropical forests are widespread across a broad range of soil properties. Even if EcM fungi have a high capability to mobilize organic forms of soil nutrients, which are thought to predominate on nutrient-depleted soils, high variability in how much of the nutrients are ultimately transferred to their host plants can disrupt the expected association between the dominance of EcM-associated trees and soil fertility. Specifically, the exchange of C and nutrients between EcM hosts and fungi is not universally reciprocal. Some EcM fungi hoard nutrients, leading to a reduction in nutrient transfer to host trees, despite receiving carbon from them<sup>38</sup> (also refs. 39,40). Under low soil nutrient availability, both AM<sup>41</sup> and EcM<sup>40</sup> mycorrhizal fungi may become nutrient limited<sup>42</sup>, thereby restricting nutrient transfer to their host. One potential strategy to counteract fungal nutrient hoarding could be the prevalence of non-mycorrhizal plants or the rejection of the fungal infection by mycotrophic plants, particularly under conditions of extremely low



**Fig. 2 | The association between the relative abundance of EcM trees and soil properties across 30 lowland tropical forests from three major tropical regions. a, b.** The y axes indicate the plot-level proportion of BA contributed by EcM trees or the relative abundance (rel. abund.) of EcM trees in BA. The x axes indicate PC1 in **a** and PC2 in **b**. The colours yellow, blue and red correspond to tropical Asia, tropical Africa and the neotropics, respectively. Dots indicate individual observations. Light-coloured lines correspond to 200 draws from the posterior predictive distribution of a model that tests the relationship between the relative abundance of EcM trees and the PCA axes. The darker-coloured lines indicate the mean predictions from these draws. PC1 was correlated with soil nutrient availability, whereas PC2 was correlated with soil texture and P availability (Fig. 1b). Dashed lines indicate that the slopes are not different from zero (Extended Data Table 4).

nutrient availability, an aspect that warrants further investigation. The non-reciprocal exchange of C and nutrient between mycorrhizal hosts and fungi adds to the complexity of the symbiotic relationship.

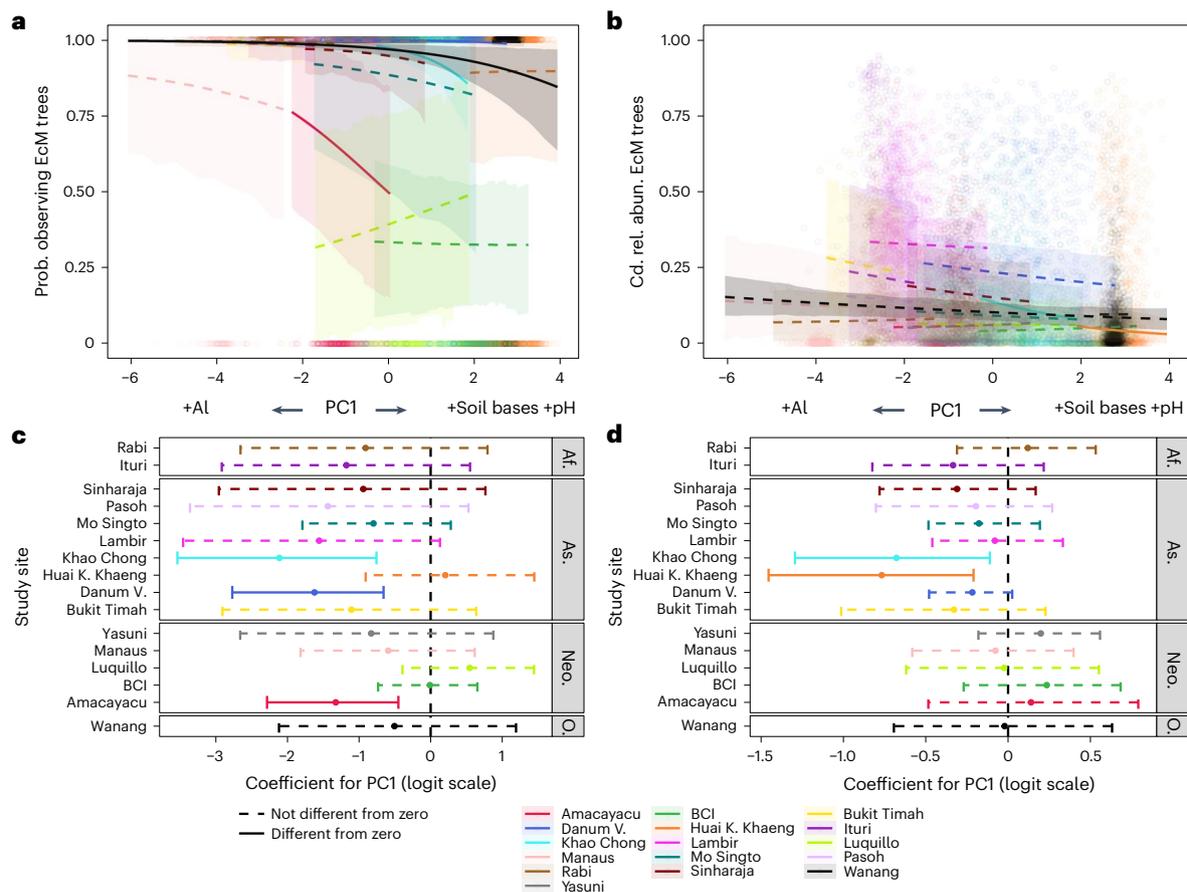
High variability in the exchange of resources between EcM symbiotic partners could result from the high phylogenetic diversity of EcM fungi<sup>43,44</sup>. EcM fungi exhibit remarkable diversity, comprising over 20,000 species from about 80 EcM fungal lineages<sup>44</sup>. In contrast, AM fungi descend from the phylum Glomeromycota (a proposed systematics places AM fungi in the Glomeromycotina subphylum within the phylum Mucoromycota<sup>45</sup>) and comprises about 345 species (as of January 2023; [www.amf-phylogeny.com](http://www.amf-phylogeny.com)). However, although there has been considerable research on AM fungi due to their widespread distribution and established association with the majority of land plants, studies on EcM fungi have been relatively limited and predominantly centred on fungi (and hosts) from the Northern Hemisphere<sup>46</sup>. In addition, there is a widespread tendency to contrast EcM and AM fungal species as monolithic entities, disregarding their inherent diversity and variations within each group. In fact, not all evolutionary lineages of EcM fungi have retained the genetic ability to degrade organic matter<sup>47</sup>, EcM fungi commonly occur in both low- and high-fertility soils independent of host identity and host distribution<sup>48</sup>, and variations in the EcM fungal composition are linked with large differences in growth rates of the host partner<sup>49</sup>. These observations, coupled with others that

indicate that even different isolates of the same EcM fungal species may have different traits and affect their host and environment in distinct ways<sup>50</sup> suggest the existence of a high variability in the functional biology of EcM fungi and that this variability could be linked to different edaphic conditions<sup>51</sup>.

The widespread distribution of EcM host lineages implicitly supports large variability in the functional biology of EcM fungi. Large well-known EcM host lineages (for example, Myrtaceae (*Eucalyptus*), Dipterocarpaceae, Fagaceae, Fabaceae (Detarioideae) and so on) occur across a very wide range of soil and hydrological environmental conditions<sup>52</sup>, supporting the idea that EcM fungi can occur across a wide range of environmental conditions (also ref. 53), probably with different costs and benefits to the host plant<sup>49</sup>. These EcM fungi could encompass different genotypes and species with different functions<sup>54</sup> or the same EcM fungal genotypes and species with plastic responses to variation in their biophysical environment<sup>55,56</sup>. Given the potential for various functional biologies within mycorrhizal fungal guilds and the importance of mycorrhizae for the dynamics of natural ecosystems, it is imperative that we improve our understanding of the functional biology of mycorrhizae, particularly in lowland tropical forests, which are still poorly understood in comparison to northern temperate and boreal regions.

We also found that the relative abundance of EcM trees is lower in lowland neotropical forests than in lowland tropical forests in Africa and Asia (Fig. 2a,b and Extended Data Table 4). Historical biogeography is an important factor explaining the low relative abundance of EcM trees in lowland neotropical forests<sup>52,57</sup>. In neotropical forests, most identified EcM hosts belong to non-dominant taxa (small trees, shrubs and lianas) within the genera *Coccoloba* (Polygonaceae), *Gnetum* (Gnetaceae) and *Guapira*, *Pisonia* and *Neea* (Nyctaginaceae)<sup>58</sup>. There are known exceptions within the Polygonaceae family (for example, *Coccoloba wifera* and *Gymnopodium floribundum*) that form monodominant patches of vegetation in their native ranges<sup>52</sup>. Dominant EcM hosts in lowland neotropical forests belong to the Fabaceae (at least four species within the genus *Aldina* and at least three species within the genus *Dicymbe*) and Dipterocarpaceae (at least one species, *Pseudomonotes tropenbosii*)<sup>57</sup>. Contrary to lowland neotropical forests, the pool of confirmed EcM-associated tree species (and individual tree sizes) is larger and appears to have a wide geographic distribution in the palaeotropics<sup>52,57</sup>. These differences in the pool of available EcM tree host species and their biogeographical distribution among tropical regions may explain the lower relative abundance of EcM trees in lowland neotropical forests. Whereas biogeographic differences have already been recognized in the literature about the current distribution of EcM plant species<sup>52</sup>, several questions remain unanswered. For instance, key knowledge gaps include why dipterocarps prevail in Asia but not in other tropical regions or why detarioid legumes do not predominate in Asia. The answers to these questions are elusive and outside the scope of this study. Better botanical data on species' taxonomies and geographic distributions, combined with corresponding phylogeographies, would help inform these knowledge gaps. Concomitantly, several thousand (about 9,000) tree species are yet to be discovered, with 40% of them estimated to be in South America<sup>59</sup>. This incomplete understanding of tree biodiversity hampers our ability to create accurate and detailed maps of mycorrhizal types in tropical regions. Acquiring comprehensive baseline information is essential for addressing this limitation.

Our results have important implications for both the mycorrhizal-associated nutrient economy framework and the current efforts to incorporate mycorrhizal nutrient acquisition into dynamic global vegetation models (DGVMs). Whereas the mycorrhizal-associated nutrient economy framework classifies temperate forests based on mycorrhizal associations and nutrient economies<sup>9</sup>, our findings reveal that the mycorrhizal associations in lowland tropical forests are far more complex and diverse than previously recognized. In lowland tropical forests, we



**Fig. 3 | The association between the probability of observing EcM trees and their relative abundance in relation to soil fertility across and within 16 lowland tropical forests. a,b.** The black lines indicate the mean quadrat-level ( $20 \times 20$  m) predictions for the probability (prob.) of observing EcM trees (a) and their conditional relative abundance (cd. rel. abun.) in BA across sites (b). Coloured dots indicate observations, varying by site (Extended Data Table 2), whereas the coloured lines show site-level mean predictions. Shaded areas around the mean lines show the 95% credible intervals of these predictions. **c,d.** Dots indicate mean site-level coefficients for the probability of observing EcM trees (c) and the relative abundance of these EcM trees (d) in relation to soil fertility (PC1). The x axes indicate the value of the coefficient on the logit scale,

and the y axes indicate the study site. Error bars indicate the 95% credible interval of the coefficient. Mean prediction lines (a,b), mean site-level coefficients (dots; c,d) and credible intervals (shaded areas and error bars) were estimated using 200 draws from the posterior predictive distribution of the Zero-Altered Beta (ZABE) regression used to estimate the probability of observing EcM trees and their conditional relative abundance in BA. Colours represent 16 sites from the lowland tropical regions of Africa (Af., two sites), Asia (As., eight sites), the neotropics (Neo., five sites) and Oceania (O., one site). Dashed lines indicate that the predicted slopes (a,b) and estimated coefficients (c,d) are not different from zero. PC1 is positively correlated with soil nutrient availability (Fig. 1c).

have identified substantial variability in the abundance of EcM trees within and across three major tropical regions, independent of soil variation. This challenges the assumption of clear gradients in nutrient economies in the transition from AM-dominated to EcM-dominated stands in lowland tropical forests, thereby questioning the applicability of the mycorrhizal-associated nutrient economy framework in these ecosystems. Differences in resource exchange among mycorrhizal partners, stemming from diverse evolutionary origins of mycorrhizal fungi, may decouple soil fertility from the advantage provided by mycorrhizal associations. Studies describing the functional biology of mycorrhizal symbiosis across a greater number of mycorrhizal and plant lineages are required, especially in the lowland tropics, where our current conception of the symbioses may be based on overinterpreted results (ref. 60) and biased by assuming that they function similarly to those in temperate and boreal regions. Furthermore, integrating a mycorrhizal framework into DGVMs to improve representations of the tropical biome under future climate change scenarios poses substantial challenges due to the limitations of models assuming a universal prevalence of AM plants in lowland tropical forests, failing to capture the observed variations in mycorrhizal dominance. Particularly in

tropical Africa and Asia, where EcM are prevalent, these limitations may become evident. Overall, our study underscores the complexity of mycorrhizal associations, raises questions about existing frameworks and highlights the difficulties in incorporating mycorrhizal dynamics into DGVMs for accurate representation of the tropical biome.

## Methods

### Tree census, estimation of BA and mycorrhizal type classification

The ForestGEO plots<sup>32</sup> are divided into  $20 \times 20$  m quadrats, the number of which varies between plots due to differences in plot sizes (from 2 ha in Bukit Timah, Singapore, to 52 ha in Lambir, Malaysia). For each ForestGEO plot, all trees  $\geq 1$  cm in DBH are tagged, mapped, measured and identified to species, following standardized census protocols<sup>32</sup>. Tree species were assigned to a mycorrhizal type (AM or EcM) based on a recently compiled database of mycorrhizal types for plant genera<sup>61</sup>. The use of checklists to assign mycorrhizal traits to host plants has been widely criticized and extensively discussed elsewhere<sup>62–64</sup>. One of the criticisms is that these lists may include errors resulting from misidentification of root mycorrhizal structures or data derived using

flawed diagnostic criteria. However, the checklist employed in this study addresses these limitations by comparing records of mycorrhizal status with expert opinions (ref. 61), thereby providing a more reliable list of plant mycorrhizal associations. For each quadrat, we calculated the proportion of BA contributed by EcM trees by dividing the BA ( $\text{m}^2$ ) of EcM trees by the total tree BA ( $\text{m}^2$ ). Multiple plant species are associated with both mycorrhizal fungi and N-fixing bacteria<sup>61,65</sup>. In our dataset, the quadrat-level number of individuals and BA of these plant species is generally less than 1% and were excluded from our analyses.

### Soil sampling and measurement of soil properties

The standard soil sampling method involved taking one sample at the centre of each  $40 \times 40$  m quadrat and another sample 2, 8 or 20 m in a randomly chosen direction to capture fine-scale variation in soil properties<sup>66</sup>. For the 2-ha plot at Bukit Timah, we collected one soil sample in every other  $20 \times 20$  m quadrat. Samples were taken from the top 10 cm, which contains many fine roots and integrates nutrient cycles. We measured exchangeable soil cations (Al, Ca, Fe, K, Mg, Na, Fe, Mn), plant-available P and pH on soils that had been air dried at ambient temperature and sieved at 2 mm. Soil pH was measured using a glass electrode in a 1:2 soil:solution ratio in water. Exchangeable soil cations were measured by extraction in 0.1 M  $\text{BaCl}_2$  (2 h, 1:30 soil to solution ratio), with detection by inductively coupled plasma optical-emission spectrometry on an Optima 7300 DV (Perkin-Elmer)<sup>67</sup>, except for Al, Mn and Fe for Ituri, Democratic Republic of Congo, which were quantified using Mehlich-III extracting solution<sup>68</sup>. TEBs were calculated as the sum of Ca, K, Mg and Na. Effective cation exchange capacity (CEC) was calculated as the sum of Al, Ca, Fe, K, Mg, Mn and Na. Percent base saturation was calculated as  $(\text{TEB} \div \text{CEC}) \times 100$ . Plant-available P was extracted in Bray-1 solution<sup>69</sup>, with detection by automated molybdate colorimetry on a Lachat Quikchem 8500 (Hach), except for BCI, Panama, where P was quantified using Mehlich-III extractant<sup>70</sup>. The Mehlich-III extracting solution is used as an alternative to the Bray-1 P (for P) and  $\text{BaCl}_2$  (for the base cations) extractants but gives relatively different concentrations depending on the acidity of the soil<sup>71</sup>. Previous studies have shown that Mehlich-III extraction yields higher results for Fe and Mn compared with other extraction methods in alkaline soils, such as  $\text{BaCl}_2$ . This has been attributed to the higher acidity of Mehlich-III and its superior acidic buffering capacity, which enhances the solubility of Fe and Mn<sup>72</sup>. For the Ituri plot, where the mean pH is 4.03, indicating an extremely acidic soil, we did not anticipate large differences between the extraction methods. Similar studies have shown that Mehlich-III extracts more P than Bray-1 under acidic conditions<sup>71</sup> (but refer to ref. 73). Because the pH in the BCI plot, Panama, is moderately acidic (mean = 5.79), we did not anticipate large differences. Moreover, we also conducted the analyses of the present manuscript excluding both the Ituri and BCI plots, and the results remained consistent. Soil texture (the proportion of sand, clay and silt content) was estimated using the sieving soil analysis technique.

### Imputation of missing soil values for the soil data at a coarse scale

Due to a variety of logistical considerations, several sites in the dataset at a coarse scale had missing values (NAs) for some soil variables (Supplementary Table 1). We imputed those NAs in the compiled dataset using a regularized iterative PCA algorithm<sup>74</sup> and then constructed the PCA on the complete dataset. This procedure involved three steps. We first selected a fixed number of dimensions via the function *estim\_ncpPCA* from the R package FactoMineR (version 2.4<sup>75</sup>) and the leave-one-out cross-validation method. The optimal number of fixed dimensions for this dataset was three. We then implemented the regularized iterative PCA algorithm with the function *imputePCA* from the R package missMDA (version 1.18<sup>74</sup>) using the fixed number of dimensions estimated in the previous step. The regularized iterative PCA algorithm

delivered the complete dataset by replacing the missing values, or NAs, with the (regularized) fitted values. Lastly, we constructed the PCA using the complete dataset via the function *PCA* implemented in the R package FactoMineR. The implemented gap-filling method did not have a direct impact on the results. For each observation in the compiled dataset, we added a constant value (one), transformed it using the natural logarithm and standardized it by calculating the z-scores before implementing the iterative PCA algorithm.

### Interpolation of soil variables for the analysis at a fine scale

For the analysis at a fine scale, we used ordinary kriging implemented in the R package geoR (version 1.8-17<sup>6</sup>) to obtain spatial predictions (spatial interpolation) for each soil variable at a  $20 \times 20$  m spatial resolution. We transformed the soil variables using a Box–Cox power transformation with a lambda value of 0, 0.5 or 1<sup>66</sup>. Assuming isotropy, we performed a polynomial trend-surface regression of the form  $s \approx x + y + x^2 + y^2 + x*y$ , where  $s$  is the transformed soil variable,  $x$  and  $y$  are the coordinates in metres of each sampling location and  $x*y$  indicate the multiplication between  $x$  and  $y$  coordinates. We extracted the residuals from the trend-surface regression and used them to compute empirical variograms with the function *variog*. We fit a set of five models (Gaussian, circular, exponential, spherical and Cauchy) to the empirical variograms to estimate covariance parameters using the function *variofit*. We selected the best model by the principle of least squares and estimated kriged means using ordinary kriging via the function *krige.conv*. We added back the kriged means to the polynomial trend and back transformed the resulting spatially predicted soil variables to the original scale using the inverse Box–Cox transform. We constructed a PCA using the pooled  $20 \times 20$  m kriged soil data across all plots to derive orthogonal composite variables to characterize the variability in soil chemical properties across and within plots with the *PCA* function in the FactoMineR package. For each kriged soil variable, we added a constant value (one), transformed it using the natural logarithm and standardized it by calculating the z-scores before constructing the PCA.

### Statistical analyses for the coarse-scale data

For the analysis at a coarse scale, we modelled the mean relative abundance in BA of EcM trees as a function of the first and second PCs (Fig. 1b; PC1 and PC2, explaining 48.2% and 26.7% of the total variation, respectively) of the PCA constructed using mean-plot-level soil data to test their association across sites from three major tropical regions (or continent) (Africa (five sites), Asia (16 sites) and the neotropics (nine sites)). We included a discrete main effect for the tropical region and its interaction with both PCs. The discrete main effect assesses whether there are variations in the mean relative abundance of EcM trees across different regions. The interaction term evaluates whether the relationship between the relative abundance of EcM trees and soil attributes, as indicated by both PCs, varies among the different regions. We excluded one site (Wanang (tag ee); Extended Data Fig. 1) from the analysis at a coarse scale because it was the only site from Oceania in our dataset. By excluding Wanang, we were able to test for interactions between the region and the PCs. The response variable (the mean relative abundance in BA of EcM trees) is continuous and restricted to the open unit interval (0,1) (greater than 0 and less than 1; Supplementary Fig. 1a). Given the nature of the data (continuous-based proportions), we constructed a GLM assuming a beta error structure. We parameterized the beta error structure using a mean ( $\mu$ ) and a precision ( $\phi$ ) parameter (beta ( $\mu, \phi$ )) instead of the more common parameterization that includes two positive shape parameters<sup>77</sup>.  $\mu$  is the mean of the response variable (the relative abundance of EcM trees) and is modelled through a logit link function.  $\phi$  is the precision, which is modelled through a log-link function, and for a fixed  $\mu$ , the larger the value of  $\phi$ , the smaller the variance of the response variable<sup>77</sup>. Both  $\mu$  and  $\phi$  can be modelled as a function of covariates under this parameterization of the beta distribution<sup>77</sup>. Accordingly, we also allowed  $\phi$  to vary by region.

### Statistical analyses for the fine-scale data

For the analysis at a fine scale, we modelled the mean relative abundance in BA of EcM trees as a function of the first (Fig. 1c, PC1, 51.4% of total variation) and second PCs (PC2, 18.6% of total variation) of the PCA constructed using fine-scale-level soil data to assess its association with soil nutrient availability among plots and within each plot. The observation unit is the quadrat (20 × 20 m), and multiple quadrats are nested within each of the 16 plots. The response variable (the mean relative abundance on BA of EcM trees) is continuous, it takes values from the open unit interval (0, 1), and it has a large probability mass at zero (quadrats without EcM trees; Supplementary Fig. 1b). We used a GLMM, assuming a Zero-Altered Beta distribution (ZABE)<sup>33,34</sup>. The ZABE distribution is a piecewise distribution or joint model with two components. A discrete component uses a Bernoulli distribution to assess if an event occurs, and a continuous component uses a beta distribution to assess the event's intensity given that it occurs<sup>33,34</sup>. The ZABE distribution has three parameters:  $\theta$  for the discrete component (or Bernoulli process) and  $\mu$  and  $\phi$  for the continuous component (or beta process).  $\theta$  is the probability that  $y$  is one, that is,  $P(y=1) = \theta$ , and  $\mu$  and  $\phi$  follow the previous description of the beta distribution, that is, beta ( $\mu, \phi$ ). The Bernoulli distribution is viewed as the distribution for EcM occurrence in a quadrat. The beta distribution is considered for the relative abundance of EcM trees in the same quadrat, conditional on the presence of EcM trees. Both  $\theta$  and  $\mu$  were modelled through a logit link function and  $\phi$  through a log-link function. It should be noted that each component of the ZABE model (discrete and continuous component) can include the same or a different set of covariates and random effects<sup>34,78</sup>.

In the model, we included the main effects of both PCs (PC1 and PC2, Fig. 1c) in both components of the ZABE model to provide a community-level (fixed effects) indication of the association of soil nutrient availability as indicated by both PCs with the probability of observing EcM trees (discrete component) and with their relative abundance (continuous component). We included site (forest plot identity) as a random intercept for both components of the ZABE model. This random intercept allowed the probability of observing EcM trees and their relative abundance to vary among plots. We included site-level random slopes for both PCs in both components of the ZABE model. This was done to allow the community-level coefficients (fixed effects) of the covariates (both PCs) to vary among plots. It also provided a within-plot indication of the association of soil nutrient availability with the probability of observing EcM trees and with their relative abundance. Local variations in topography are known to impact soil properties and community composition<sup>79</sup>. We calculated quadrat-level slope and convexity (the mean elevation of one 20 × 20 m quadrat relative (minus) to the mean of its immediate neighbours) from elevation data using the function *fgeo\_topography* from the *fgeo* package (version 1.1.4<sup>80</sup>). We included these two covariates and the quadrat-level elevation in both components of the ZABE model to characterize the terrain among (as fixed effects) and within forests (as random slopes). We did not include mean annual air temperature and mean annual precipitation in the model because the former was relatively constant across plots, and their within-plot variation among quadrats was probably minimal. The total quadrat-level BA was highly variable within and among plots (Extended Data Fig. 2a), indicating high variation in forest structural complexity, successional stages and forest maturity among and within plots<sup>81</sup>. We included quadrat-level total BA as an additional covariate in both processes of the ZABE model to account for this variability. This addition also helped to control for differences in exposure, as it accounts for a higher opportunity of observing EcM trees and a higher relative abundance of EcM trees in quadrats with larger BAs or the opposite. Total BA was transformed with the natural logarithm and was included as both fixed and random (slope) effects. All predictors included in the model were standardized by calculating their z-score. This standardization allowed parameters

to be comparable. All  $R^2$  determined by pairwise correlations among covariates among forests were <0.4 (Supplementary Table 2), indicating that all the described covariates could safely be included in the model<sup>82</sup>.

We accounted for the spatial dependency in both components of the ZABE model by specifying a spatially structured random effect using a reparameterization of the Besag–York–Mollié (BYM2) model<sup>83</sup>. The BYM2 model in the continuous component of the ZABE model was incorporated as a copy of the BYM2 model in the discrete component. That is, the spatial random effect in the continuous component (copy) shares the same hyperparameters as the discrete component (original) but is multiplied by a scale parameter  $\beta$  estimated from the data<sup>84</sup>. The incorporated copy feature also links both components of the ZABE model.

We performed an additional analysis for the data at a fine scale to test whether our results are robust to differences in the spatial scale (there might be just a few emergent trees in a 20 × 20 m quadrat) by using a resolution (quadrat size) of 40 × 40 m instead of the 20 × 20 m resolution. Using a resolution of 40 × 40 m implied suppressing an entire row and column from the plots matrices.

### Statistical software and model evaluation and inference

We performed all the analyses in R (version 4.2.0<sup>85</sup>). For the analysis at a coarse scale, we fitted the beta model in Stan<sup>86</sup>, which fits models using the Hamiltonian Monte Carlo algorithm, with its interface to R via cmdstanr (version 0.4.0<sup>87</sup>) and using the package brms (version 2.16.1<sup>88,89</sup>). We estimated the model using four chains of 2,000 iterations, each with a burn-in fraction of 1/2. We monitored Markov Chain mixing properties and checked parameter convergence graphically via trace plots of the estimated coefficients and by checking the Rhat metric<sup>90</sup>. For the analysis at a fine scale, we fitted the ZABE model (GLMM) using INLA (version 22.05.07<sup>91</sup>) in R because of its speed and the straightforward implementation of spatial random effects. A description of the priors used in both analyses is described in Supplementary Note 1. We inspected the goodness-of-fit of the full model for the analyses at a coarse (Supplementary Fig. 2a) and fine scale (Supplementary Fig. 2bc) via posterior predictive model checks<sup>92</sup>, where predictions from the fitted model were compared to the observed data. Results are presented based on the mean and 95% credible intervals indicated in square brackets.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

ForestGEO plot data can be obtained upon request via the ForestGEO portal at <http://ctfs.si.edu/datarequest/>. All data sources are listed in Extended Data Table 1. PCA axes and the contribution (proportion) of EcM trees to basal area can be found at <https://doi.org/10.5281/zenodo.10044772> ref. 93.

### Code availability

The code to run the analyses at both coarse and fine scales can be found at <https://doi.org/10.5281/zenodo.10044772> ref. 93.

### References

1. Lambers, H., Mougél, C., Jaillard, B. & Hinsinger, P. Plant–microbe–soil interactions in the rhizosphere: an evolutionary perspective. *Plant Soil* **321**, 83–115 (2009).
2. Smith, S. E. & Read, D. J. *Mycorrhizal Symbiosis* (Academic Press, 2008).
3. Tedersoo, L. & Bahram, M. Mycorrhizal types differ in ecophysiology and alter plant nutrition and soil processes. *Biol. Rev.* **94**, 1857–1880 (2019).

4. Branco, S. et al. Mechanisms of stress tolerance and their effects on the ecology and evolution of mycorrhizal fungi. *New Phytol.* **235**, 2158–2175 (2022).
5. Jiang, Y. et al. Plants transfer lipids to sustain colonization by mutualistic mycorrhizal and parasitic fungi. *Science* **356**, 1172–1175 (2017).
6. Howard, N. et al. The potential role of Mucoromycotina ‘fine root endophytes’ in plant nitrogen nutrition. *Physiol. Plant.* **174**, e13715 (2022).
7. Hoysted, G. A. et al. Direct nitrogen, phosphorus and carbon exchanges between Mucoromycotina ‘fine root endophyte’ fungi and a flowering plant in novel monoxenic cultures. *New Phytol.* **238**, 70–79 (2023).
8. Brundrett, M. C. & Tedersoo, L. Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytol.* **220**, 1108–1115 (2018).
9. Phillips, R. P., Brzostek, E. & Midgley, M. G. The mycorrhizal-associated nutrient economy: a new framework for predicting carbon–nutrient couplings in temperate forests. *New Phytol.* **199**, 41–51 (2013).
10. Wurzburger, N. et al. Mycorrhizal fungi as drivers and modulators of terrestrial ecosystem processes. *New Phytol.* **213**, 996–999 (2017).
11. Soudzilovskaia, N. A. et al. Global mycorrhizal plant distribution linked to terrestrial carbon stocks. *Nat. Commun.* **10**, 5077 (2019).
12. Gadgil, R. L. & Gadgil, P. D. Mycorrhiza and litter decomposition. *Nature* **233**, 133–133 (1971).
13. Rozmoš, M. et al. Organic nitrogen utilisation by an arbuscular mycorrhizal fungus is mediated by specific soil bacteria and a protist. *ISME J.* **16**, 676–685 (2022).
14. Wang, L. et al. A core microbiome in the hyphosphere of arbuscular mycorrhizal fungi has functional significance in organic phosphorus mineralization. *New Phytol.* **238**, 859–873 (2023).
15. Tedersoo, L., Bahram, M. & Zobel, M. How mycorrhizal associations drive plant population and community biology. *Science* **367**, eaba1223 (2020).
16. Becquer, A. et al. in *Advances in Botanical Research* (ed. Cánovas, F. M.) 77–126 (Academic Press, 2019).
17. Averill, C. et al. Global imprint of mycorrhizal fungi on whole-plant nutrient economics. *Proc. Natl Acad. Sci. USA* **116**, 23163–23168 (2019).
18. Seyfried, G. S., Dalling, J. W. & Yang, W. H. Mycorrhizal type effects on leaf litter decomposition depend on litter quality and environmental context. *Biogeochemistry* **155**, 21–38 (2021).
19. Steidinger, B. S. et al. Climatic controls of decomposition drive the global biogeography of forest–tree symbioses. *Nature* **569**, 404–408 (2019).
20. Read, D. J. Mycorrhizas in ecosystems. *Experientia* **47**, 376–391 (1991).
21. Braghiere, R. K. et al. Mycorrhizal distributions impact global patterns of carbon and nutrient cycling. *Geophys. Res. Lett.* **48**, e2021GL094514 (2021).
22. Barceló, M. et al. Climate drives the spatial distribution of mycorrhizal host plants in terrestrial ecosystems. *J. Ecol.* **107**, 2564–2573 (2019).
23. Lokonda, M. et al. Are soils under monodominant *Gilbertiodendron dewevrei* and under adjacent mixed forests similar? A case study in the Democratic Republic of Congo. *J. Trop. Ecol.* **34**, 176–185 (2018).
24. Lee, H. S. et al. Floristic and structural diversity of mixed dipterocarp forest in Lambir Hills National Park, Sarawak, Malaysia. *J. Trop. Sci.* **14**, 379–400 (2002).
25. Newbery, D. M. et al. Transient dominance in a central African rain forest. *Ecol. Monogr.* **83**, 339–382 (2013).
26. Beard, J. S. The Mora forests of Trinidad, British West Indies. *J. Ecol.* **33**, 173–192 (1946).
27. Steidinger, B. S. et al. Variability in potential to exploit different soil organic phosphorus compounds among tropical montane tree species. *Funct. Ecol.* **29**, 121–130 (2015).
28. Koele, N. et al. No globally consistent effect of ectomycorrhizal status on foliar traits. *New Phytol.* **196**, 845–852 (2012).
29. Godoy, R. & Marín, C. in *Mycorrhizal Fungi in South America* (eds Pagano, M. C. & Lugo, M. A.) 315–341 (Springer International Publishing, 2019).
30. Marín, C. et al. Geological history and forest mycorrhizal dominance effects on soil fungal diversity in Chilean temperate rainforests. *J. Soil Sci. Plant Nutr.* **23**, 734–745 (2023).
31. Weemstra, M. et al. Lithological constraints on resource economies shape the mycorrhizal composition of a Bornean rain forest. *New Phytol.* **228**, 253–268 (2020).
32. Davies, S. J. et al. ForestGEO: understanding forest diversity and dynamics through a global observatory network. *Biol. Conserv.* **253**, 108907 (2021).
33. Liu, F. & Eugenio, E. C. A review and comparison of Bayesian and likelihood-based inferences in beta regression and zero-or-one-inflated beta regression. *Stat. Methods Med. Res.* **27**, 1024–1044 (2018).
34. Zuur, A. F. & Ieno, E. N. *Beginner’s Guide to Spatial, Temporal and Spatial-temporal Ecological Data Analysis with R-INLA: GAM and Zero-inflated Models* (Highland Statistics Limited, 2018).
35. Barceló, M. et al. Mycorrhizal tree impacts on topsoil biogeochemical properties in tropical forests. *J. Ecol.* **110**, 1271–1282 (2022).
36. Newbery, D. M., Alexander, I. J. & Rother, J. A. Phosphorus dynamics in a lowland African rainforest: the influence of ectomycorrhizal trees. *Ecol. Monogr.* **67**, 367–409 (1997).
37. Henkel, T. W. Monodominance in the ectomycorrhizal *Dicymbe corymbosa* (Caesalpinaceae) from Guyana. *J. Trop. Ecol.* **19**, 417–437 (2003).
38. Hasselquist, N. J. et al. Greater carbon allocation to mycorrhizal fungi reduces tree nitrogen uptake in a boreal forest. *Ecology* **97**, 1012–1022 (2016).
39. Franklin, O. et al. Forests trapped in nitrogen limitation – an ecological market perspective on ectomycorrhizal symbiosis. *New Phytol.* **203**, 657–666 (2014).
40. Näsholm, T. et al. Are ectomycorrhizal fungi alleviating or aggravating nitrogen limitation of tree growth in boreal forests? *New Phytol.* **198**, 214–221 (2013).
41. Treseder, K. K. & Allen, M. F. Direct nitrogen and phosphorus limitation of arbuscular mycorrhizal fungi: a model and field test. *New Phytol.* **155**, 507–515 (2002).
42. Allen, M. F. *Mycorrhizal Dynamics in Ecological Systems* (Cambridge Univ. Press, 2022).
43. Martin, F., Kohler, A., Murat, C., Veneault-Fourrey, C. & Hibbett, D. S. Unearthing the roots of ectomycorrhizal symbioses. *Nat. Rev. Microbiol.* **14**, 760–773 (2016).
44. Tedersoo, L. & Smith, M. E. in *Biogeography of Mycorrhizal Symbiosis* (ed. Tedersoo, L.) 125–142 (Springer International Publishing, 2017).
45. Spatafora, J. W. et al. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* **108**, 1028–1046 (2016).
46. Dickie, I. A. & Moyersoen, B. Towards a global view of ectomycorrhizal ecology. *New Phytol.* **180**, 263–265 (2008).
47. Pellitier, P. T. & Zak, D. R. Ectomycorrhizal fungi and the enzymatic liberation of nitrogen from soil organic matter: why evolutionary history matters. *New Phytol.* **217**, 68–73 (2018).
48. Peay, K. G. et al. Lack of host specificity leads to independent assortment of dipterocarps and ectomycorrhizal fungi across a soil fertility gradient. *Ecol. Lett.* **18**, 807–816 (2015).

49. Anthony, M. A. et al. Forest tree growth is linked to mycorrhizal fungal composition and function across Europe. *ISME J.* **16**, 1327–1336 (2022).
50. Plett, K. L. et al. Intra-species genetic variability drives carbon metabolism and symbiotic host interactions in the ectomycorrhizal fungus *Pisolithus microcarpus*. *Environ. Microbiol.* **23**, 2004–2020 (2021).
51. Corrales, A. et al. Variation in ectomycorrhizal fungal communities associated with *Oreomunnea mexicana* (Juglandaceae) in a Neotropical montane forest. *Mycorrhiza* **26**, 1–17 (2015).
52. Tedersoo, L. in *Biogeography of Mycorrhizal Symbiosis* (ed. Tedersoo, L.) 469–531 (Springer International Publishing, 2017).
53. Corrales, A. et al. Diversity and distribution of tropical ectomycorrhizal fungi. *Mycologia* **114**, 919–933 (2022).
54. Pena, R. & Polle, A. Attributing functions to ectomycorrhizal fungal identities in assemblages for nitrogen acquisition under stress. *ISME J.* **8**, 321–330 (2014).
55. Hazard, C. et al. Contrasting effects of intra- and interspecific identity and richness of ectomycorrhizal fungi on host plants, nutrient retention and multifunctionality. *New Phytol.* **213**, 852–863 (2017).
56. Hortal, S. et al. Role of plant–fungal nutrient trading and host control in determining the competitive success of ectomycorrhizal fungi. *ISME J.* **11**, 2666–2676 (2017).
57. Corrales, A., Henkel, T. W. & Smith, M. E. Ectomycorrhizal associations in the tropics—biogeography, diversity patterns and ecosystem roles. *New Phytol.* **220**, 1076–1091 (2018).
58. Nouhra, E. R. et al. in *Mycorrhizal Fungi in South America* (eds Pagano, M. C. & Lugo, M. A.) 73–95 (Springer International Publishing, 2019).
59. Cazzolla Gatti, R. et al. The number of tree species on Earth. *Proc. Natl Acad. Sci. USA* **119**, e2115329119 (2022).
60. Karst, J., Jones, M. D. & Hoeksema, J. D. Positive citation bias and overinterpreted results lead to misinformation on common mycorrhizal networks in forests. *Nat. Ecol. Evol.* **7**, 501–511 (2023).
61. Soudzilovskaia, N. A. et al. FungalRoot: global online database of plant mycorrhizal associations. *New Phytol.* **227**, 955–966 (2020).
62. Brundrett, M. & Tedersoo, L. Misdiagnosis of mycorrhizas and inappropriate recycling of data can lead to false conclusions. *New Phytol.* **221**, 18–24 (2019).
63. Bueno, G. et al. Conceptual differences lead to divergent trait estimates in empirical and taxonomic approaches to plant mycorrhizal trait assignment. *Mycorrhiza* **29**, 1–11 (2019).
64. Bueno, C. G. et al. Misdiagnosis and uncritical use of plant mycorrhizal data are not the only elephants in the room: a response to Brundrett & Tedersoo (2019) ‘Misdiagnosis of mycorrhizas and inappropriate recycling of data can lead to false conclusions’. *New Phytol.* **224**, 1415–1418 (2019).
65. Tedersoo, L. et al. Global database of plants with root-symbiotic nitrogen fixation: NodDB. *J. Veg. Sci.* **29**, 560–568 (2018).
66. John, R. et al. Soil nutrients influence spatial distributions of tropical tree species. *Proc. Natl Acad. Sci. USA* **104**, 864–869 (2007).
67. Hendershot, W. H., Lalonde, H. & Duquette, M. in *Soil Sampling and Methods of Analysis* (eds Carter, M. R. & Gregorich, E. G.) 197–206 (CRC Press, 2008).
68. Mehlich, A. Mehlich 3 soil test extractant: a modification of Mehlich 2 extractant. *Commun. Soil Sci. Plant Anal.* **15**, 1409–1416 (1984).
69. Bray, R. H. & Kurtz, L. T. Determination of total, organic, and available forms of phosphorus in soils. *Soil Sci.* **59**, 39–46 (1945).
70. Wolf, J. A. et al. Geospatial observations on tropical forest surface soil chemistry. *Ecology* **96**, 2313–2313 (2015).
71. Fukuda, M. et al. Evaluation of the Mehlich 3 reagent as an extractant for cations and available phosphorus for soils in Mozambique. *Commun. Soil Sci. Plant Anal.* **48**, 1462–1472 (2017).
72. Bibiso, M. et al. Evaluation of universal extractants for determination of selected micronutrients from soil. *Bull. Chem. Soc. Ethiop.* **29**, 199–213 (2015).
73. Tran, T. S. et al. Evaluation of Mehlich-III extractant to estimate the available P in Quebec soils. *Commun. Soil Sci. Plant Anal.* **21**, 1–28 (1990).
74. Josse, J. & Husson, F. Handling missing values in exploratory multivariate data analysis methods. *J. Soci.été Fr. Stat.* **153**, 79–99 (2012).
75. Lê, S., Josse, J. & Husson, F. FactoMineR: an R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).
76. Ribeiro, P. J. Jr et al. geoR: analysis of geostatistical data. *R. package version 1*, 8–1 (2020).
77. Ferrari, S. & Cribari-Neto, F. Beta regression for modelling rates and proportions. *J. Appl. Stat.* **31**, 799–815 (2004).
78. Blangiardo, M. et al. Spatial and spatio-temporal models with R-INLA. *Spat. Spatiotemporal Epidemiol.* **4**, 33–49 (2013).
79. Valencia, R. et al. Tree species distributions and local habitat variation in the Amazon: large forest plot in eastern Ecuador. *J. Ecol.* **92**, 214–229 (2004).
80. Lepore, M. et al. Fgeo: analyze forest diversity and dynamics. R package version 1.1.4 (2019).
81. Peña-Claros, M. Changes in forest structure and species composition during secondary forest succession in the Bolivian Amazon. *Biotropica* **35**, 450–461 (2003).
82. Dormann, C. F. et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27–46 (2013).
83. Riebler, A. et al. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Stat. Methods Med. Res.* **25**, 1145–1165 (2016).
84. Gómez-Rubio, V. *Bayesian Inference with INLA* (CRC Press, 2020).
85. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2023).
86. Carpenter, B. et al. Stan: a probabilistic programming language. *J. Stat. Softw.* **76**, jss.v076.i01 (2017).
87. Gabry, J. & Češnovar, R. Cmdstanr: R interface to ‘CmdStan’. R package version 0.4.0 (2021).
88. Bürkner, P.-C. Brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, jss.v080.i01 (2017).
89. Bürkner, P.-C. Advanced Bayesian multilevel modeling with the R package brms. *R. J.* **10**, 395–411 (2018).
90. Gelman, A. et al. *Bayesian Data Analysis* (CRC press, 2013).
91. Rue, H., Martino, S. & Chopin, N. Approximate Bayesian inference for latent gaussian models using integrated nested laplace approximations (with discussion). *J. R. Stat. Soc. B* **71**, 319–392 (2009).
92. Gabry, J. et al. Visualization in Bayesian workflow. *J. R. Stat. Soc. A* **182**, 389–402 (2019).
93. Medina-Vega, J. A. et al. Dataset and code accompanying the study by Medina-Vega et al. in *Nature Ecology & Evolution*: tropical tree ectomycorrhiza are distributed independently of soil nutrients. Zenodo <https://doi.org/10.5281/zenodo.10044772> (2023).
94. Quesada, C. A. et al. Variations in chemical and physical properties of Amazon forest soils in relation to their genesis. *Biogeosciences* **7**, 1515–1541 (2010).
95. Chave, J. et al. Above-ground biomass and productivity in a rain forest of eastern South. *Am. J. Trop. Ecol.* **24**, 355–366 (2008).
96. Fanin, N. et al. Does variability in litter quality determine soil microbial respiration in an Amazonian rainforest? *Soil Biol. Biochem.* **43**, 1014–1022 (2011).
97. Libalah, M. B. et al. Shift in functional traits along soil fertility gradient reflects non-random community assembly in a tropical African rainforest. *Plant Ecol. Evol.* **150**, 265–278 (2017).

98. Nkongolo, N. V., Mbuyi, J. J. K. & Lokonda, M. W. Quantification of soil carbon in Ituri forest, Democratic Republic of Congo. In *Proc. Global Symposium on Soil Organic Carbon* 151–153 (Food and Agriculture Organization of the United Nations, 2017).
99. Baillie, I. C. et al. Stoichiometry of cationic nutrients in Phaeozems derived from skarn and acrisols from other parent materials in lowland forests of Thailand. *Geoderma Reg.* **12**, 1–9 (2018).
100. Sukri, R. et al. Habitat associations and community structure of dipterocarps in response to environment and soil conditions in Brunei Darussalam, Northwest Borneo. *Biotropica* **44**, 595–605 (2012).
101. Ross, S. M. & Dykes, A. in *Tropical Rainforest Research—Current Issues: Conf. Proc* (eds Edwards, D. S. et al.) 259–270 (Springer Netherlands, 1996).
102. Dent, D. H. et al. Nutrient fluxes via litterfall and leaf litter decomposition vary across a gradient of soil nutrient supply in a lowland tropical rain forest. *Plant Soil* **288**, 197–215 (2006).
103. Anderson-Teixeira, K. J. et al. CTFs-ForestGEO: a worldwide network monitoring forests in an era of global change. *Glob. Change Biol.* **21**, 528–549 (2015).
104. Moraga, P. *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny* (CRC Press, 2019).

## Acknowledgements

We express our gratitude to the dedicated field and data technicians whose meticulous data-gathering efforts were indispensable to this research. Their pivotal contributions formed the backbone of our study. Our gratitude also extends to the teams of scientists behind the papers and datasets that enriched our primary dataset. This research and J.A.M.-V. were supported as part of the Next Generation Ecosystem Experiments-Tropics, funded by the US Department of Energy, Office of Science, Office of Biological and Environmental Research. For site-specific acknowledgements, please refer to Supplementary Table 3.

## Author contributions

J.A.M.-V. and S.J.D. conceptualized the study, coordinated the data compilations, designed the analysis and interpreted the data. J.A.M.-V. performed data analyses. J.A.M.-V. led the writing of the paper with

inputs from S.J.D., D.F.R.P.B., J.W.D., S.E.R. and D.Z. S.A., A.A., P.B., W.Y.B., S.B., N.C., J.C., A.A.d.O., Á.D., S.E., C.E.N.E., J.F., S.P.H., A.I., S.K., S.K.Y.L., J.-R.M., H.M., D.M., M.B.M., A.N., R.N., N.V.N., V.N., M.J.O., R.P., N.P., G.R., S.T., J.T., M.U., R.V., A.V., T.L.Y., J.K.Z. and D.Z. provided coordination and leadership, data management and quality control, travel, consumables and commented on the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41559-023-02298-0>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41559-023-02298-0>.

**Correspondence and requests for materials** should be addressed to José A. Medina-Vega.

**Peer review information** *Nature Ecology & Evolution* thanks Leho Tedersoo, César Marín and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

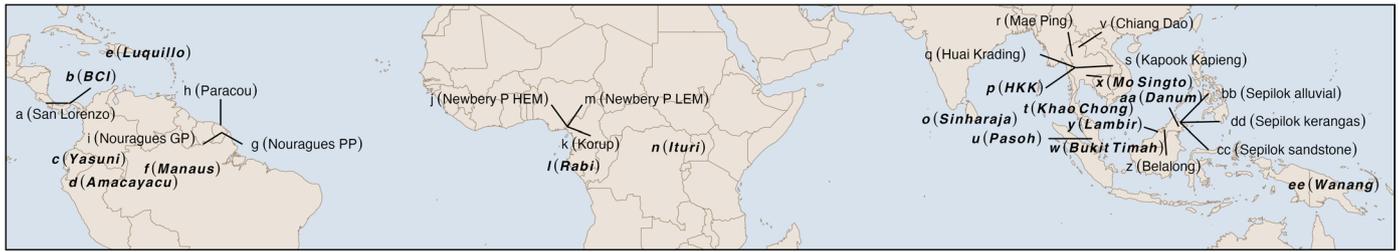
Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2024

**José A. Medina-Vega**<sup>1</sup>✉, **Daniel Zuleta**<sup>1</sup>, **Salomón Aguilar**<sup>2</sup>, **Alfonso Alonso**<sup>3</sup>, **Pulchérie Bissiengou**<sup>4</sup>, **Warren Y. Brockelman**<sup>5,6</sup>, **Sarayudh Bunyavejchewin**<sup>7</sup>, **David F. R. P. Burslem**<sup>8</sup>, **Nicolás Castaño**<sup>9</sup>, **Jérôme Chave**<sup>10</sup>, **James W. Dalling**<sup>2,11</sup>, **Alexandre A. de Oliveira**<sup>12</sup>, **Álvaro Duque**<sup>13</sup>, **Sisira Ediriweera**<sup>14</sup>, **Corneille E. N. Ewango**<sup>15</sup>, **Jonah Filip**<sup>16</sup>, **Stephen P. Hubbell**<sup>17</sup>, **Akira Itoh**<sup>18</sup>, **Somboon Kiratiprayoon**<sup>19</sup>, **Shawn K. Y. Lum**<sup>20</sup>, **Jean-Remy Makana**<sup>15</sup>, **Hervé Memiaghe**<sup>21</sup>, **David Mitre**<sup>2</sup>, **Mohizah Bt. Mohamad**<sup>22</sup>, **Anuttara Nathalang**<sup>5</sup>, **Reuben Nilus**<sup>23</sup>, **Nsalambi V. Nkongolo**<sup>24,25</sup>, **Vojtech Novotny**<sup>26,27</sup>, **Michael J. O'Brien**<sup>28</sup>, **Rolando Pérez**<sup>2</sup>, **Nantachai Pongpattananurak**<sup>7</sup>, **Glen Reynolds**<sup>29</sup>, **Sabrina E. Russo**<sup>30,31</sup>, **Sylvester Tan**<sup>22</sup>, **Jill Thompson**<sup>32</sup>, **María Uriarte**<sup>33</sup>, **Renato Valencia**<sup>34</sup>, **Alberto Vicentini**<sup>35</sup>, **Tze Leong Yao**<sup>36</sup>, **Jess K. Zimmerman**<sup>37</sup> & **Stuart J. Davies**<sup>1</sup>

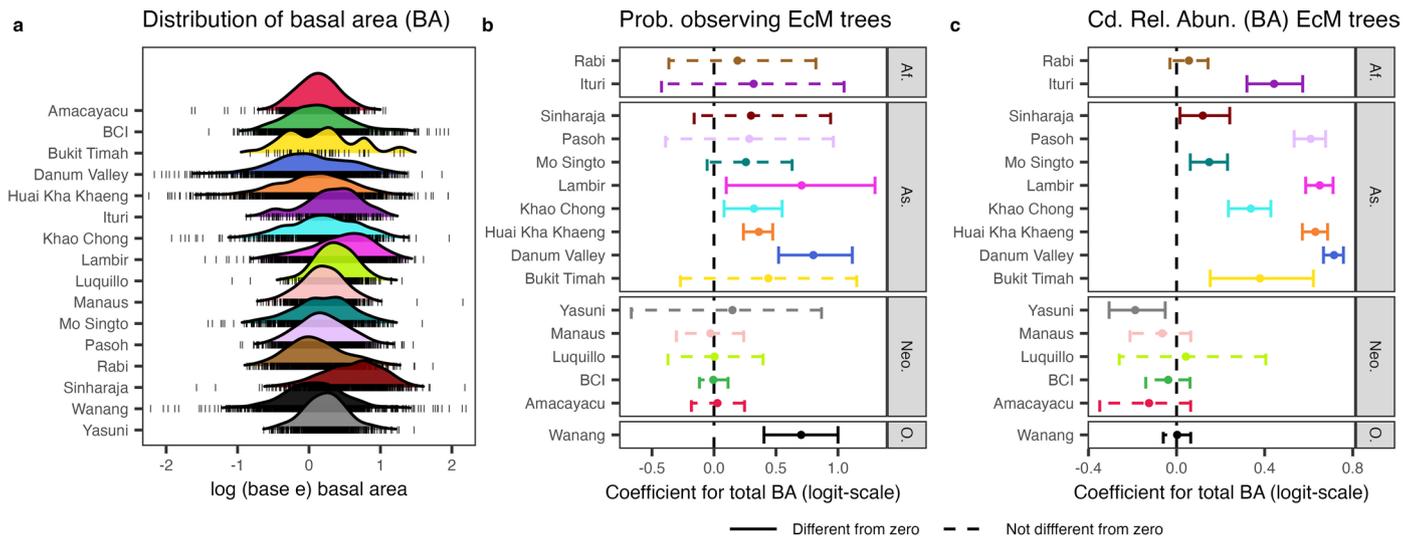
<sup>1</sup>Forest Global Earth Observatory, Smithsonian Tropical Research Institute, Washington, DC, USA. <sup>2</sup>Smithsonian Tropical Research Institute, Balboa, Panama. <sup>3</sup>Center for Conservation and Sustainability, Smithsonian National Zoo and Conservation Biology Institute, Washington, DC, USA. <sup>4</sup>Herbier National du Gabon, Institut de Pharmacopée et de Médecine Traditionnelle, Libreville, Gabon. <sup>5</sup>National Biobank of Thailand, National Science and Technology Development Agency, Khlong Luang, Thailand. <sup>6</sup>Institute of Molecular Biosciences, Mahidol University, Nakhon Pathom, Thailand. <sup>7</sup>Thai Long-Term Forest Ecological Research Project, Department of Forest Biology, Faculty of Forestry, Kasetsart University, Bangkok, Thailand. <sup>8</sup>School of Biological Sciences, University of Aberdeen, Aberdeen, UK. <sup>9</sup>Herbario Amazónico Colombiano, Instituto Amazónico de Investigaciones Científicas Sinchi, Bogotá, Colombia. <sup>10</sup>Laboratoire Evolution et Diversité Biologique, CNRS, UPS, IRD, Université Paul Sabatier, Toulouse, France. <sup>11</sup>Department of Plant Biology, University of Illinois Urbana-Champaign, Urbana, IL, USA. <sup>12</sup>Departamento de Ecologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil. <sup>13</sup>Departamento de Ciencias Forestales, Universidad Nacional de Colombia Sede Medellín, Medellín, Colombia. <sup>14</sup>Department of Science and Technology, Uva Wellasa University, Badulla, Sri Lanka. <sup>15</sup>Faculty of Sciences, University of Kisangani, Kisangani, Democratic Republic of the Congo. <sup>16</sup>Binatang Research Center, Madang, Papua New Guinea. <sup>17</sup>Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA. <sup>18</sup>Graduate School of Science, Osaka Metropolitan University, Osaka, Japan. <sup>19</sup>Faculty of Science and Technology, Thammasat University (Rangsit),

Pathum Thani, Thailand. <sup>20</sup>Asian School of the Environment, Nanyang Technological University, Singapore, Singapore. <sup>21</sup>Institut de Recherche en Ecologie Tropicale, Centre National de la Recherche Scientifique et Technologique, Libreville, Gabon. <sup>22</sup>Sarawak Forestry Department, Kuching, Malaysia. <sup>23</sup>Sabah Forestry Department, Forest Research Centre, Sandakan, Malaysia. <sup>24</sup>School of Science, Navajo Technical University, Crownpoint, NM, USA. <sup>25</sup>Institut Facultaire des Sciences Agronomiques (IFA) de Yangambi, Kisangani, Democratic Republic of the Congo. <sup>26</sup>Biology Centre, Institute of Entomology of the Czech Academy of Sciences, Ceske Budejovice, Czech Republic. <sup>27</sup>Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic. <sup>28</sup>Estación Experimental de Zonas Áridas, Consejo Superior de Investigaciones Científicas, Almería, Spain. <sup>29</sup>Southeast Asia Rainforest Research Partnership (SEARRP), Kota Kinabalu, Malaysia. <sup>30</sup>School of Biological Sciences, University of Nebraska, Lincoln, NE, USA. <sup>31</sup>Center for Plant Science Innovation, University of Nebraska, Lincoln, NE, USA. <sup>32</sup>UK Centre for Ecology and Hydrology, Edinburgh, UK. <sup>33</sup>Department of Ecology, Evolution, and Environmental Biology, Columbia University, New York, NY, USA. <sup>34</sup>Escuela de Ciencias Biológicas, Pontificia Universidad Católica del Ecuador, Quito, Ecuador. <sup>35</sup>Coordenação de Dinâmica Ambiental (CODAM), Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, Brazil. <sup>36</sup>Forestry and Environment Division, Forest Research Institute Malaysia, Kepong, Malaysia. <sup>37</sup>Department of Environmental Sciences, University of Puerto Rico, San Juan, PR, USA. ✉e-mail: [jamedinavega@gmail.com](mailto:jamedinavega@gmail.com)



**Extended Data Fig. 1 | Location of the 31 study sites.** Letters indicate the tags used to identify plots in the principal component analysis (PCA) of soil data, constructed using coarse scale soil data (Fig. 1b and Extended Data Table 1).

The names of each study site are enclosed in parentheses. Thirty sites were used in the analysis at a coarse scale, whereas 16 sites (shown in bold and italics) were used for the fine scale analysis (see Methods). Map made with Natural Earth.



**Extended Data Fig. 2 | Variation in quadrat-level basal area and its association with the probability of observing EcM trees and their relative abundance in 16 lowland tropical forests.** **a**, Shows the distribution of the quadrat-level total basal area (BA) after applying a natural logarithm transformation. The x-axis represents the transformed total BA for each  $20 \times 20$  m quadrat, whereas the y-axis indicates the study site. Vertical lines at the base of each density curve indicate individual observations. **b, c**, Present mean site-level coefficients, with panel b representing the probability (Prob.) of observing EcM trees and panel c for their relative abundance (conditional on the presence of EcM trees; Cd. Rel. Abun.) in relation to the quadrat-level

total BA. The x-axes show the value of the coefficient on the logit scale, with the y-axes again showing the study site. Error bars show the 95% credible interval of the coefficient. These coefficients and their credible intervals derive from 200 draws from the Zero-Altered Beta (ZABE) regression's posterior predictive distribution. This regression estimated the probability of observing EcM trees and their conditional relative abundance in BA, with the total quadrat-level basal area being logarithmically transformed before the analyses. The study includes 16 sites from lowland tropical regions in Africa (Af., two sites), Asia (As., eight sites), the neotropics (Neo., five sites), and Oceania (O., one site). Dashed lines indicate that the coefficients are not different from zero.

Extended Data Table 1 | List of the 31 study sites<sup>94–102</sup>

Tag	Region	Site	Tree data	DBH cut-off	Soil data
a	Neotropics	San Lorenzo	ForestGEO	>= 1 cm	ForestGEO
b	Neotropics	BCI	ForestGEO	>= 1 cm	ForestGEO
c	Neotropics	Yasuni	ForestGEO	>= 1 cm	ForestGEO
d	Neotropics	Amacayacu	ForestGEO	>= 1 cm	ForestGEO
e	Neotropics	Luquillo	ForestGEO	>= 1 cm	ForestGEO
f	Neotropics	Manaus	ForestGEO	>= 1 cm	ForestGEO, 94
g	Neotropics	Nouragues PP	ForestGEO 95	>= 10 cm	ForestGEO
h	Neotropics	Paracou	ForestGEO	>= 10 cm	96
i	Neotropics	Nouragues GP	ForestGEO 95	>= 10 cm	ForestGEO
j	Africa	Newbery P-HEM †	36	>= 10 cm	25,36
k	Africa	Korup	ForestGEO	>= 1 cm	97
l	Africa	Rabi	ForestGEO	>= 1 cm	ForestGEO
m	Africa	Newbery P-LEM †	36	>= 10 cm	36
n	Africa	Ituri (Lenda 1)	ForestGEO	>= 1 cm	98, ForestGEO
o	Asia	Sinharaja	ForestGEO	>= 1 cm	ForestGEO
p	Asia	HKK (Huai Kha Khaeng)	ForestGEO	>= 1 cm	ForestGEO
q	Asia	Huai Krading	ForestGEO	>= 1 cm	99
r	Asia	Mae Ping	ForestGEO	>= 1 cm	99
s	Asia	Kapook Kapieng	ForestGEO	>= 1 cm	99
t	Asia	Khao Chong	ForestGEO	>= 1 cm	ForestGEO
u	Asia	Pasoh	ForestGEO	>= 1 cm	ForestGEO
v	Asia	Chiang Dao	ForestGEO	>= 1 cm	99
w	Asia	Bukit Timah	ForestGEO	>= 1 cm	ForestGEO
x	Asia	Mo Singto	ForestGEO	>= 1 cm	ForestGEO
y	Asia	Lambir	ForestGEO	>= 1 cm	ForestGEO
z	Asia	Belalong	ForestGEO	>= 1 cm	100,101
aa	Asia	Danum	ForestGEO	>= 1 cm	ForestGEO
bb	Asia	Sepilok alluvial	ForestGEO	>= 5 cm	102
cc	Asia	Sepilok sandstone	ForestGEO	>= 5 cm	102
dd	Asia	Sepilok kerangas	ForestGEO	>= 5 cm	102
ee*	Oceania	Wanang	ForestGEO	>= 1 cm	ForestGEO

Tag is an identification ID for each site (see Extended Data Fig. 1). Region indicates the major tropical region. Site indicates the site name. Tree data and soil data specify the data sources for the tree and soil data for each forest, respectively. DBH cut-off indicates the minimum DBH that was sampled. † For the Newbery P-HEM (tag j) and P-LEM (tag m) sites, we used the mean (proportion) contribution of EcM trees to basal area, and the mean soil data from both the P-HEM and P-LEM transects, as reported in<sup>36</sup>. \* The sole site from Oceania (Wanang, tag ee) was excluded from the coarse scale analysis due to sample size (see Methods, statistical analyses for the coarse-scale data).

## Extended Data Table 2 | List of the 16 study sites for the analysis at a fine scale

Tag	Region	Site	Elev-m	MAT-C	MAP-mm	N. dry months	Size (ha)	Long.	Lat.	N.	N. Zeros
d	Neotropics	Amacayacu, Colombia	94	25.8	3215	0	25	-70.2678	-3.8091	625	298
b	Neotropics	Barro Colorado Island (BCI), Panamá	120	27.1	2551	4	50	-79.8461	9.1543	1250	769
w	Asia	Bukit Timah, Singapore	99	26.9	2473	0	2	103.78	1.35	50	1
aa	Asia	Danum Valley, Malaysia	150	26.7	2822	0	50	117.688	5.1019	1250	24
p	Asia	Huai Kha Khaeng, Thailand	596	23.5	1476	5	50	99.217	15.6324	1250	282
n	Africa	Ituri, D.R. Congo	775	24.3	1682	3	10	28.5826	1.4368	250	0
t	Asia	Khao Chong, Thailand	235	27.1	2611	2.5	24	99.798	7.5435	600	57
y	Asia	Lambir, Malaysia	174	26.6	2664	0	52	114.017	4.1865	1300	3
e	Neotropics	Luquillo, Puerto Rico	381	22.8	3548	0	16	-65.816	18.3262	400	303
f	Neotropics	Manaus, Brazil	60	26.7	2600	2	25	-59.7858	-2.4417	625	120
x	Asia	Mo Singto, Thailand	770	23.5	2100	5.5	30	101.35	14.4333	750	30
u	Asia	Pasoh, Malaysia	80	27.9	1788	0	50	102.313	2.982	1250	0
l	Africa	Rabi, Gabon	41	26	2299	3-4	25	9.88	-1.9246	625	2
o	Asia	Sinharaja, Sri Lanka	500	22.5	5016	0	25	80.4023	6.4023	625	6
ee	Oceania	Wanang, Papua New Guinea	140	25.8	4000	0	50	145.267	-5.25	1250	21
c	Neotropics	Yasuni, Ecuador	230	28.3	3081	0	25	-76.397	-0.6859	625	2

Tag is an identification ID for each site (see Extended Data Fig. 1). Region indicates the major tropical region (neotropics [five sites], Asia [eight sites], Africa [two sites], and Oceania [one site]). Site indicates the site and country name. Elev-m is the mean elevation in meters. MAT-C is the mean annual temperature in Celsius. MAP-mm is the mean annual precipitation in mm. N. dry months is the mean number of dry months in a year. Size (ha) indicates the area in hectares. Longitude and Latitude give the coordinates of each site. N. is the number of quadrats (20 × 20 m) per site and N. Zeros indicates the number of quadrats where no EcM individuals were found. Biophysical data were obtained from<sup>32,103</sup>.

Extended Data Table 3 | Principal component analyses (PCA) of soil data at coarse and fine scales

Soil attribute	PC1	PC2
<b>Coarse scale soil data</b>		
% Clay	0.04 (-)	<b>22.79 (+)</b>
% Silt	3.43 (+)	0.96 (+)
% Sand	0.35 (-)	<b>23.56 (-)</b>
P (mg.kg-1)	3.86 (+)	<b>8.38 (-)</b>
Soil pH in water	<b>10.04 (+)</b>	1.15 (-)
Mg (cmolc.kg-1)	<b>13.46 (+)</b>	0.04 (+)
K (cmolc.kg-1)	<b>10.53 (+)</b>	3.44 (-)
Ca (cmolc.kg-1)	<b>14.35 (+)</b>	0.01 (+)
Na (cmolc.kg-1)	0.22 (+)	<b>15.68 (+)</b>
Mn (cmolc.kg-1)	3.41 (+)	<b>18.19 (+)</b>
Al (cmolc.kg-1)	5.52 (-)	<b>4.72 (+)</b>
CEC (cmolc.kg-1)	<b>9.93 (+)</b>	0.67 (+)
TEB (cmolc.kg-1)	<b>14.5 (+)</b>	0.15 (+)
% BS	<b>10.35 (+)</b>	0.27 (+)
Variance (%)	48.18	26.66
Cum. variance (%)	48.18	74.84
Eigenvalue	6.72	3.71
<b>Fine scale soil data</b>		
Al (mg.kg-1)	<b>14.71 (-)</b>	1.57 (+)
Ca (mg.kg-1)	<b>19.22 (+)</b>	3.69 (+)
Fe (mg.kg-1)	7.98 (-)	<b>13.03 (+)</b>
K (mg.kg-1)	<b>15.01 (+)</b>	1.61 (+)
Mg (mg.kg-1)	<b>18.40 (+)</b>	4.04 (+)
Mn (mg.kg-1)	3.01 (+)	<b>11.07 (+)</b>
Na (mg.kg-1)	0.38 (-)	<b>44.68 (+)</b>
P (mg.kg-1)	2.42 (+)	<b>20.17 (-)</b>
Soil pH in water	<b>18.86 (+)</b>	0.14 (-)
Variance (%)	51.39	18.61
Cum. variance (%)	51.39	70
Eigenvalue	4.62	1.67

We conducted two separate PCA analyses on soil attributes for lowland tropical forests. The 'Coarse scale soil data' analysis used 14 soil attributes from 30 forests, while the 'Fine scale soil data' used 9 soil attributes from 16 forests. For both analyses, we calculated the variable contribution (%), eigenvalues, variance, and cumulative explained variance for the first two components (PC1 and PC2). In bold are the contributions (%) whose |loadings| (loadings are equal to the coordinates of the variables divided by the square root of the eigenvalue associated with the component) were greater than the mean, indicating the most important contributions. Symbols in parentheses indicate the direction of the association between the soil variable and the PCs.

**Extended Data Table 4 | Coefficients for the analysis of the relative abundance of EcM trees in relation to soil properties at a coarse scale**

Parameter	Estimate	LCI	UCI
<b>Mean - <math>\mu</math></b>			
Intercept (Region Africa)	-1.63	-3.39	0.21
Region Asia	1.01	-0.83	2.82
Region Neotropics	<b>-2.53</b>	<b>-4.53</b>	<b>-0.54</b>
PC1 (Region Africa)	-0.37	-1.27	0.63
PC2 (Region Africa)	0.35	-0.74	1.42
Region Asia:PC1	0.31	-0.72	1.24
Region Neotropics:PC1	0.47	-0.58	1.4
Region Asia:PC2	-0.22	-1.35	0.93
Region Neotropics:PC2	-0.39	-1.47	0.75
<b>Precision - <math>\phi</math></b>			
Intercept	1.03	-0.23	2.05
Region Asia	0.4	-0.86	1.8
Region Neotropics	2.46	-0.16	4.49

Parameter is the variable included in the model. Estimate, LCI and UCI are the mean estimate, the lower and the upper 95% credible intervals, respectively, calculated from the posterior predictive distribution. Coefficients with credible intervals that contain zero are considered to have negligible influence, whereas coefficients in bold exclude zero and are considered influential. The reference level for the categorical variable 'Region' is Africa.

**Extended Data Table 5 | Coefficients for the analysis of the relative abundance of EcM trees among and within 16 lowland tropical forests at a fine scale**

Parameter	Estimate	LCI	UCI	Parameter	Estimate	LCI	UCI
<b>Fixed effect</b>	<b>Discrete component - Bernoulli process</b>			<b>Random effect</b>	<b>Discrete component - Bernoulli process</b>		
Intercept	<b>3.48</b>	<b>2.77</b>	<b>4.19</b>	$\tau$ site	0.19	0.09	0.43
PC1	<b>-0.88</b>	<b>-1.44</b>	<b>-0.33</b>	$\tau$ site PC1	1.31	0.59	2.58
PC2	-0.12	-0.46	0.22	$\tau$ site PC2	3.98	1.39	8.22
Convexity	0.15	-0.001	0.3	$\tau$ site Convexity	21.01	10.39	37.3
Slope	-0.11	-0.33	0.11	$\tau$ site Slope	5.48	2.97	8.95
Elevation	0.27	-0.7	1.26	$\tau$ site Elevation	0.83	0.51	1.27
(log) BA	<b>0.3</b>	<b>0.14</b>	<b>0.45</b>	$\tau$ site (log) BA	8.42	5.15	12.54
	<b>Continuous component - beta process</b>				<b>Continuous component - beta process</b>		
Intercept	<b>-2.16</b>	<b>-2.52</b>	<b>-1.81</b>	Precision - $\phi$	6.17	5.9	6.5
PC1	-0.16	-0.33	0.01	$\tau$ site	1.25	0.66	2.05
PC2	-0.08	-0.18	0.01	$\tau$ site PC1	5.69	2.9	9.93
Convexity	0.01	-0.02	0.04	$\tau$ site PC2	35.82	8.29	110.76
Slope	-0.06	-0.11	0.00	$\tau$ site Convexity	1018.54	172.13	3382.4
Elevation	0.17	-0.26	0.6	$\tau$ site Slope	67.91	30.81	142.4
(log) BA	<b>0.23</b>	<b>0.2</b>	<b>0.27</b>	$\tau$ site Elevation	1.11	0.5	2.11
				$\tau$ site (log) BA	6.99	3.85	12.84
					<b>Spatial random terms (BYM2)</b>		
					<b>Discrete component - Bernoulli process</b>		
				$\tau$ ID.z	2.69	2.09	3.34
				$\phi$ ID.z	0.96	0.94	0.97
					<b>Continuous component - beta process</b>		
				$\beta$ ID.o	0.79	0.73	0.84

Parameter is the variable included in the model, and its estimate, LCI, and UCI (95% credible intervals) are calculated from the posterior predictive distribution. The table is partitioned both horizontally and vertically. Horizontally, the left section shows fixed effect coefficients, whereas the right section shows the random effect coefficients. Vertically, the table is split into three sections. The top and middle sections show coefficients for the discrete and continuous components of the ZABE model, respectively. The bottom section lists coefficients for the spatial random terms (BYM2 model) for both components. Fixed effects coefficients in bold exclude zero and are considered influential, whereas fixed effects coefficients with credible intervals that contain zero (not in bold) are considered negligible. Random effects are restricted to be positive. Precision is indicated by  $\tau$ , with  $\tau = 1/\sigma$ , where  $\sigma$  is the standard deviation. For the spatial random terms in the discrete component, the parameter ' $\tau$  ID.z' indicates the posterior of the precision of the spatial field and the parameter ' $\phi$  ID.z' indicates the posterior of the precision of the mixing parameter (see<sup>104</sup> for details). The parameter ' $\beta$  ID.o' in the spatial random terms for the continuous component represents the estimated scale parameter. All continuous predictors in the model were standardized using their z-score, and the total basal area was transformed using the natural logarithm before standardization ((log) BA).

**Extended Data Table 6 | Coefficients for the analysis of the relative abundance of EcM trees among and within 13 lowland tropical forests at a fine scale**

Parameter	Estimate	LCI	UCI
<b>Discrete component - Bernoulli process</b>			
Intercept	3.69	2.89	4.5
PC1	-0.37	-1.07	0.32
PC2	-0.18	-0.63	0.27
Convexity	0.18	-0.03	0.38
Slope	-0.11	-0.4	0.17
Elevation	-0.04	-1.12	1.04
(log) BA	<b>0.25</b>	<b>0.06</b>	<b>0.45</b>
<b>Continuous component - beta process</b>			
Intercept	-2.17	-2.56	-1.78
PC1	-0.15	-0.35	0.05
PC2	-0.1	-0.22	0.02
Convexity	0.01	-0.03	0.04
Slope	-0.06	-0.12	0
Elevation	0.39	-0.08	0.85
(log) BA	<b>0.21</b>	<b>0.17</b>	<b>0.26</b>

Parameter is the variable included in the model, and its estimate, LCI, and UCI (95% credible intervals) are calculated from the posterior predictive distribution. The table is split into two sections. The top and bottom sections show coefficients for the discrete and continuous components of the ZABE model, respectively. Coefficients in bold exclude zero and are considered influential, whereas coefficients with credible intervals that contain zero (not in bold) are considered negligible. All continuous predictors in the model were standardized using their z-score, and the total basal area was transformed using the natural logarithm before standardization ((log) BA). This analysis excludes three forests: Amacayacu in Colombia, Khao Chong in Thailand and Danum Valley in Malaysia. Random effects are not shown.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

ForestGEO plot data can be obtained upon request via the ForestGEO portal at <http://cfs.si.edu/datarequest/>. All data sources are listed in Extended Data Table 1. PCA axes and the contribution (proportion) of EcM trees to basal area can be found at <https://doi.org/10.5281/zenodo.10044772>

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

n/a

Reporting on race, ethnicity, or other socially relevant groupings

n/a

Population characteristics

n/a

Recruitment

n/a

Ethics oversight

n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

We compiled data on the relative abundance of EcM trees (i.e., the proportion of basal area [BA] contributed by EcM trees) and soil properties of 31 lowland tropical forests from the Forest Global Earth Observatory (ForestGEO) plot network of research sites and the literature. We created two datasets: one dataset at a coarse scale using mean plot level data for 30 sites, and another dataset at a fine scale using spatially explicit (20 x 20m quadrats) data from 16 sites. The fine scale dataset included data from 16 sites, because these sites had all trees  $\geq 1$  cm in DBH identified and the most complete and consistently measured set of soil variables. We constructed two PCAs using the soil data: one at a coarse scale and one at a fine scale. We selected the first two principal components of those PCAs to characterize variation in soil properties at each scale (i.e., fine and coarse scale). We then used these selected PCA axes in a generalized linear model (GLM) and a generalized linear mixed effects model (GLMM) to assess the association between EcM tree abundance in basal area (BA) and soil properties. The GLM was used to analyze coarse scale associations across forests, while the GLMM was used to analyze fine scale associations within and across forests.

Research sample

There were 31 lowland tropical forests included in the study, with 29 ForestGEO sites and two sites from literature. The plot sizes ranged between 2 and 52 hectares. A dataset at a coarse scale was created using data from 30 sites, with mean plot-level information on the relative abundance of EcM trees in terms of basal area and soil properties, including soil chemistry (i.e., Al, Ca, K, Mg, Mn, Na, CEC, and TEB in  $\text{cmolc.kg}^{-1}$ , plant-available P in  $\text{mg.kg}^{-1}$ , pH, and percent base saturation [BS]) and soil texture (i.e., the proportion of sand, clay, and silt content). Another dataset was created at a fine scale using data from 16 sites, where all trees  $\geq 1$  cm in DBH were identified to species, and with the most complete and consistently measured set of soil variables, including Al, Ca, K, Mg, Na, Fe, Mn, plant-available P, and pH. The fine scale dataset was based on 20 x 20 m quadrats.

Sampling strategy

For the 29 ForestGEO plots, all free-standing woody plants were identified to species, mapped, and measured for DBH. The DBH cut-off varied among plots, as indicated in Extended Data Table 1. The standard soil sampling method involved taking one sample at the center of each 40x40 m quadrat and another sample at a distance of 2, 8, or 20 m in a randomly chosen direction to capture fine-scale variation in soil properties. For the 2-ha Bukit Timah plot, we collected one sample in every other 20x20m quadrat. Soil samples were collected from the top 10cm. In cases where soil data were not available (i.e., sites where we did not measure collected and

measured soil samples), we obtained mean plot-level soil data from the literature. We included two sites from the literature, and these two sites were included only in the dataset at a coarse scale.

**Data collection**  
For free-standing woody plants, all tree censuses were conducted by multiple technicians during long field campaigns. The protocol used for the tree census is described in the publication "Tropical forest census plots: methods and results from Barro Colorado Island, Panama and a comparison with other plots." by Condit, R. (1998) and is currently being implemented in all ForestGEO plots.  
For soil data, multiple technicians collected soil data in the field following the same protocol (refer to the sampling strategy section in this pdf).

**Timing and spatial scale**  
We used two spatial scales. In the coarse level dataset, we used mean-plot level data. In the fine scale dataset, we used spatially explicit data (i.e., 20 x 20 m quadrat). We used one census per site.

**Data exclusions**  
We conducted two analyses, one analysis at a coarse scale using mean plot level data and another analysis at a fine scale using spatially explicit (20 x 20 m quadrats) data. For the analysis at a coarse scale, we excluded one site from Oceania because one of the goals of this study was to test for interactions between tropical region and soil fertility. As Oceania had only one observation (n = 1), we excluded it from the analysis. However, we included this site in the analysis at a fine scale as the objective of this second analysis was to assess the association between fine-scale soil fertility and the fine-scale distribution and abundance of EcM trees across and within sites.

**Reproducibility**  
No experiments have been conducted for this study. Tree census and soil data collection are described in detail and we also provide detailed description of the statistical methods that can be used to replicate the analysis in other studies.

**Randomization**  
For this study, randomization applies to soil data collection. The standard soil sampling method involved taking one sample at the center of each 40x40 m quadrat and another sample at a distance of 2, 8, or 20 m in a randomly chosen direction to capture fine-scale variation in soil properties. However, for the plot at Bukit Timah, which is only 2 hectares, one sample was collected in every 20x20m quadrat. Randomization is not relevant for the tree data.

**Blinding**  
Blinding is not relevant for the study design.

Did the study involve field work?  Yes  No

## Field work, collection and transport

**Field conditions**  
Censuses for each plot take an average of six months to a full year. Climatic data is described in Extended Data Table 2. All sites are located in the tropics, and temperature is relatively similar among plots.

**Location**  
The 31 sites are located across the tropics. Nine sites in the Neotropics. Five sites in Africa. One site in Oceania. Sixteen sites in Asia. Maps showing site locations are displayed in Fig. 1a and Extended Data Fig. 1.

**Access & import/export**  
All research activities carried out by ForestGEO and the partnering institutions at each site are conducted under strict regulations in accordance with local government laws that regulate each site, with the aim of minimizing any type of disturbance.

**Disturbance**  
Each site has a system of well defined trails to minimize impact.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input type="checkbox"/>	<input checked="" type="checkbox"/> Plants

### Methods

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Dual use research of concern

Policy information about [dual use research of concern](#)

### Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

- | No                                  | Yes   |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Public health              |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> National security          |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Crops and/or livestock     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Ecosystems                 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Any other significant area |

### Experiments of concern

Does the work involve any of these experiments of concern:

- | No                                  | Yes  |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Demonstrate how to render a vaccine ineffective                             |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Confer resistance to therapeutically useful antibiotics or antiviral agents |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Enhance the virulence of a pathogen or render a nonpathogen virulent        |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Increase transmissibility of a pathogen                                     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Alter the host range of a pathogen  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Enable evasion of diagnostic/detection modalities                           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Enable the weaponization of a biological agent or toxin                     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Any other potentially harmful combination of experiments and agents         |