Global Ecology and Biogeography

WILEY

Global Ecology and Biogeography A Journal of Macroecology

PERSPECTIVE OPEN ACCESS

Global Diversity Estimates Need to Acknowledge Species–Area Relationships

S. Faurby[1,2] | T. J. Matthews[3,4] | D. Silvestro[1,2,5,6]

[1]Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden | [2]Gothenburg Global Biodiversity Centre, Gothenburg, Sweden | [3]GEES (School of Geography, Earth and Environmental Sciences) and Birmingham Institute of Forest Research, University of Birmingham, Birmingham, UK | [4]CE3C Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group/CHANGE, Faculty of Agricultural Sciences and Environment, University of Azores, Angra do Heroísmo, Açores, Portugal | [5]Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland | [6]Swiss Institute of Bioinformatics, Basel, Switzerland

Correspondence: S. Faurby (soren.faurby@bioenv.gu.se)

Received: 30 January 2026 | **Revised:** 26 May 2026 | **Accepted:** 5 June 2026

Handling Editor: Sandra Nogue

Keywords: biased estimates | cryptic species | overestimates | SAR | undescribed species | unrecorded diversity

ABSTRACT

Background: Many studies have tried to estimate the number of undescribed species based on the known diversity. These estimates often rely on extrapolation based on data from a limited number of species. Although statistical methods provide accurate inference when generalizing from a random sample, their predictions will be biased when based on a non-random sample unless the sampling process is explicitly accounted for.

Problem: In this paper, we argue that this is a fundamental issue in many estimates of unrecorded biodiversity. We show that the sample of species used in biodiversity extrapolation represents disproportionately common and abundant taxa, which leads to an overestimation of extrapolated diversity. We discuss this issue in the context of three specific cases: estimates of plant-associated insect diversity, estimates of parasite diversity and estimates of cryptic species diversity.

Implications: For the example of plant-associated insects, we provide an estimate of the magnitude of the error, but insufficient data are currently available to estimate the magnitude of the problem for the other examples. Our findings cast doubt over previous attempts to estimate the number of undescribed species, suggesting that they provide consistent overestimations.

1 | Introduction

How many species are there on Earth? This apparently simple question has proved extremely difficult to answer. For many groups we do not even know how many species have been properly described (Mora et al. 2011), and a large fraction of the ones we have described may turn out to be synonyms (Costello et al. 2013). It is therefore not surprising that the number of unknown species is even harder to assess. Yet, this question has fascinated many scientists, with numerous estimates having been made over the last 40 years based on different methods, with several reviews summarizing these estimates and their associated limitations (e.g., May 1988; Stork et al. 2015; Wiens 2023). However, many of these estimates rely on the same problematic assumption, which appears to have been largely neglected so far: they assume that the species we have data for and that are used to derive estimates of unknown species numbers, are a random subset of all species. In biology, however, truly random samples are extremely rare.

One of the most studied patterns in biology is the species–area relationship (SAR). The SAR, a cornerstone of the equilibrium island biogeography model developed by MacArthur and Wilson (1967), describes how the size of islands (or areas more

---

generally) positively correlates with the number of species. In the classic example of endemic amphibians and reptiles on Caribbean islands, the pattern is evident (Figure 1a), with the number of species ranging from 1 on Saba (13 km^2) to 205 on Hispaniola (76,192 km^2). We use this example dataset to illustrate the problem of extrapolating total species richness from a non-random subset of measurements. If only the diversity of a subset of n islands were known, we could extrapolate to estimate the total diversity of the archipelago as the average number of species observed across the n sampled islands multiplied by the total number of islands (here, 7). If the subset of islands is selected randomly, we will on average get unbiased estimates of the true diversity, with increasing precision as we increase the number of sampled islands (Figure 1b). However, if we sampled islands non-randomly with respect to their size (in our example with probability proportional to the logarithm of the area), our extrapolation would become biased toward higher diversity (Figure 1c).

It is obvious that, as a result of the SAR, larger islands have more species and that we therefore cannot use a sampling methodology that is more likely to sample larger islands and not correct for it in our diversity estimates. In this paper we will argue that many estimates of unrecorded diversity are effectively doing the equivalent of this. This is because these analyses typically involve trying to estimate the number of species associated with other species (e.g., the number of insect species associated with individual tree species, or the number of parasite species

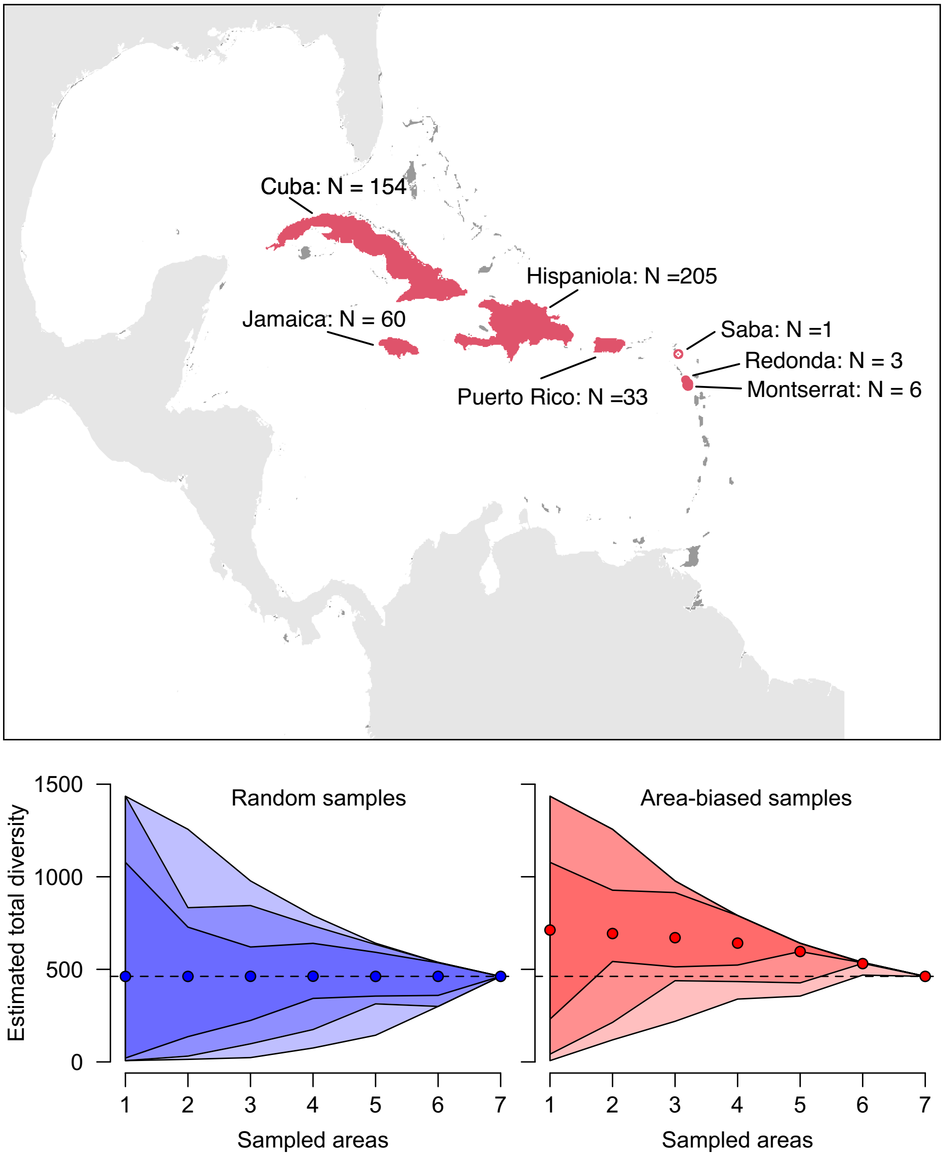

FIGURE 1 | Illustrated effects of biased sampling. (a) The richness of amphibians and reptiles on seven Caribbean islands and its relationship to island size. We re-analysed the same islands and organism groups as MacArthur and Wilson (1967) used in their seminal work, but we updated the data to (i) follow a more recent taxonomy and (ii) only include single island endemics following Caribherp (2025). In (b) we extrapolated the total diversity of the archipelago based on data from a set of 1 or more islands randomly sampled without replacement. The circles report the mean of extrapolated diversity across 10,000 random samples, with the dashed line showing the true total diversity ($N = 462$). Shaded areas show the 95%, 80% and 50% quantiles. In (c) we repeated the analysis but resampled the areas with sampling weights proportional to the logarithm of island area (in km^2). The results clearly show that biased sampling produces biased estimates of total diversity.

associated with individual host species), as these groups are predicted to contain a large proportion of unrecorded diversity. However, when undertaking analyses of this nature, the species we have data for and use to extrapolate to all species are disproportionally common and widespread. Just as larger areas can support more species, more widespread and common species likely have more associated species. This likely leads to global diversity extrapolations that are potentially substantially overestimated. The issue is likely present in many types of analyses, and here, we discuss three specific cases focused on the estimation of: the diversity of plant-associated insects based on data of insect species interacting with individual tree species, parasite diversity based on host associations, and the number of cryptic species.

A small number of studies have stressed the importance of considering SARs in richness estimation, including an interesting comparison of multiple models used to upscale diversity from plot to country level (Kunin et al. 2018). In this paper, we argue that extrapolations based on a few well-studied species, which is a common methodology employed in analyses of global biodiversity (e.g., Forbes et al. 2018; Larsen et al. 2017; Li and Wiens 2023), likewise should incorporate SARs, and that the results of published estimates that did not are likely overestimates.

We note that alternative methods exist for extrapolating richness that are not influenced by the SAR issue discussed here, but these generally have separate problems. For example, one of the most cited global diversity estimation studies utilizes a method that is based on an observed strong correlation between the number of species and the number of higher order taxonomical groupings (Mora et al. 2011), but this approach has been criticized for a lack of a theoretical basis (Scheffers et al. 2012). Other methods rely on patterns in the number of species discovered per year, but models fitted to these data normally fail to converge unless the majority of species in a group are already known (Bebber et al. 2007). Thus, they cannot be employed to estimate the unrecorded diversity of hyperdiverse but understudied groups. Furthermore, results are frequently compared between methods, with the assumption that a method is more reliable when it generates estimates that are similar to the estimates derived from other methods (see e.g., Stork et al. 2015). Under this approach, the inferred reliability of completely independent methods (such as Mora et al. 2011) may still be affected by methods that are our focus here, that is, those that should but do not properly account for SAR dynamics.

2 | Estimates Based on Plant Associated Insects

One of the earliest attempts to estimate unrecorded diversity, which is also one of the studies where the bias we discuss is easiest to see, is the estimate of global insect diversity by Erwin (1982). Under the assumption that the vast majority of all species were insects, he focused on this group and estimated the total diversity based on the approximately 1200 insect species associated with a single neotropical tree species (*Luehea seemannii*). Here, insect species associated with a tree are considered as insects for which the tree species represents a non-negligible fraction of their habitat. This will include both herbivores feeding on the tree itself as well as predators, fungivores and scavengers using the tree as habitat. Erwin then estimated the fraction of specialist insect species, that is, species only found on a single tree species, and assumed that this, as well as the number of non-specialist associated insects, was similar for all tropical tree species. Based on these estimates and assumptions, Erwin extrapolated the total global insect diversity to be around 30,000,000 species. Several studies have since argued that the fraction of host specific insects was vastly overestimated by Erwin and that most insects are specialists to tree genera rather than species and that his estimate therefore was an overestimate (Basset et al. 1996; Hamilton et al. 2010; Novotny et al. 2002). Another assumption in Erwin's study, which has received less attention, is that the insect diversity from a moderately common tree species is an unbiased estimate of the average insect diversity for a neotropical tree. This second assumption however likely also leads to a substantial overestimate of the total unknown diversity.

If we imagine conducting a similar experiment in Germany, we could for example base our diversity estimate on the 9 insect species associated with *Taxus* (Yew), the 12 species associated with *Ilex* (Holly), the 499 species associated with *Betula* (Birch) or the 728 species associated with *Salix* (Willow) (Brändle and Brandl 2002). In this regard, it is important to note that *Taxus* and *Ilex*, which have the lowest associated insect diversity, are among the rarest tree genera in Germany while *Betula* and *Salix*, with the most associated insects, may be among the most abundant (Brändle and Brandl 2002).

It has long been recognized that there is a strong relationship between plant abundance and the number of associated as well as host-specific insects (Brändle and Brandl 2002; Janzen 1970; Southwood 1960, 1961), and that this relationship can be understood in terms of the SAR (Janzen 1968). In this context, more abundant tree species represent a greater ‘area’ for insect species to inhabit, and thus more abundant trees are predicted to have more associated insect species. This is also related to the classic Janzen-Connell hypothesis, where high diversity in tropical rainforests is maintained because rarer species have a competitive advantage over common ones by escaping from specialized herbivores and parasites (Connell 1971). Southwood (1960) proposed two linked reasons for such a SAR: (i) that the number of encounters between an insect and a tree is a function of the tree species' abundance, ‘thereby influencing the rate of species evolution of the ability to feed on that tree’ and (ii) a target effect (*sensu* Lomolino 1990), where more abundant tree species have a greater likelihood of being found by immigrant insect individuals (Southwood 1960). If rarer tree species are associated with fewer specialized herbivore species, this relationship between tree abundance and associated insect diversity should be taken into account when estimating diversity.

Returning to the European example discussed above, anyone wandering through Germany picking a random tree is much more likely to select a birch or willow tree than a yew or a holly. We can further see that the same thing applies to the single tree Erwin (1982) selected. *Luehea seemannii* mainly occurs in central America (POWO 2024), and Erwin studied it in Costa Rica, although it extends into the Amazon rainforest. The Amazon rainforest contains ~15,000 tree species, of which the most abundant comprises billions of individuals while around a third of species have estimated population sizes of 1000 individuals or less (ter Steege et al. 2020). The Amazonian population of *L. seemannii*, which seems to only represent a small fraction of the entire population of the species (Condit 2019), has been estimated to number 0.45 million trees (ter Steege et al. 2020), making it roughly the 5000th most abundant out of the estimated 15,000 tree species in the Amazon rainforest.

If we generalize, it is clear that if a study is based on the species represented by a randomly selected subset of individuals, the species studied will nearly always be much more common than the median species in that assemblage when species are ranked by abundance. In this context, the species–area relationship will further interact with one or both of two other macroecological patterns: (i) the approximately log-normal distribution of species range sizes (Gaston 1998) and (ii) the species abundance distribution, which has often been shown to be well approximated by logseries distributions (i.e., most species are relatively rare) across scales, from local to global assemblages (Baldridge et al. 2016; Callaghan et al. 2023; ter Steege et al. 2020). The combined result of this is that selecting one or more species of tree, investigating their associated fauna, and then generalizing to all species is nearly guaranteed to lead to overestimates of the number of associated insects.

The logic we are using here has also partly been incorporated into recent estimates of unrecorded fungi, although this was done without noting species–area relationships and species abundance patterns. Earlier studies estimated unrecorded plant-associated fungi by assuming a global ratio of the number of fungi and plants in an area (Hawksworth 1991), whereas more recent studies have noted that the relative number of fungi per plant is lower in the tropics and higher at the site compared to the country level (Niskanen et al. 2023). Although not discussed in these studies, the results match our expectations. First, at the site level, the most common species are disproportionally likely to be present, and these should have more associated fungal species. Second, tropical regions have relatively higher numbers of rare species (Enquist et al. 2019), and rare plants should have fewer associated fungal species.

In summary, while a SAR—in terms of tree abundance and the number of associated insects—is known to operate in temperate systems (Brändle and Brandl 2002; Southwood 1961), it has not yet been incorporated into estimates of tropical insect diversity. Doing so would reduce previous estimates of undescribed tropical insects.

3 | Estimates of the Number of Animal Parasites and Parasitoids

Although much attention to global unrecorded diversity has focused on plant-associated insects, there are other extremely diverse, albeit even less understood, groups; many of these are parasites or parasitoids (Forbes et al. 2018; Larsen et al. 2017). Estimating the number of parasites is therefore an essential part of estimating total undescribed diversity. A challenge in this regard is defining what a parasite actually is. One attempt at defining this was made by Poulin and Morand (2000), who note a loose definition of a parasite as 'obligate feeding on a living organism without death to the host', and a stricter one further requiring that 'an organism must live on or in its host for a significant portion of its life to be considered a parasite'. Plant-eating insects are generally not included in discussions of parasites, and for the purpose of this paper, their inclusion or exclusion is irrelevant. We do however stress that most plant-eating insects would fit both the loose and strict criteria listed above and that the ecological dynamics between plants and insects therefore seem very similar to the dynamics between hosts and parasites.

Previous diversity estimates for parasites (e.g., Forbes et al. 2018; Larsen et al. 2017), just like previous estimates of plant-eating insects, have ignored the SAR and the parasite estimates may therefore be overestimates. Since the bias we focus on requires both the presence of a SAR and biased sampling, we will first show why a SAR is expected in the context of parasite diversity, and then highlight the biased sampling.

The existence of a SAR in parasites has been somewhat controversial but there are strong reasons to assume its presence. Although some studies have only found weak or even no support for a parasite SAR (e.g., Strona and Fattorini 2014), a global meta-analysis provided strong evidence of the pattern (Kamiya et al. 2014). The SAR for plant-eating insects is well known (Southwood 1960) and it is difficult to use theoretical arguments to explain why a SAR should apply for insects and plants but not for parasites and hosts. In addition, Strona and Fattorini (2014) highlight that sampling is difficult for parasites, meaning that the data used to search for a SAR are often sparse and noisy. In line with a SAR for parasites, it seems clear that host species with narrow ranges and small population sizes are unlikely to sustain specialist parasites. There is a strong relationship between population size or range size and extinction rate (Griffin and Drake 2008; Payne and Finnegan 2007) and since no parasite will infect all host individuals at the same time, parasite species in hosts with small populations will face high extinction risk. Parasites vary widely in the minimum host population size they require, but there are at least some species requiring very large population sizes. For instance, the host-specific measles virus that only affects humans requires population sizes of at least 500,000 humans to survive in a population (Black 1966). Population sizes are only known for a subset of species but among birds, which is one of the only groups where global population sizes have been estimated for nearly all species, the median population size at the species level was only 450,000 (Callaghan et al. 2021). Given that many species are divided into multiple populations, it seems likely that most species do not have any populations large enough to sustain parasites like measles, thus again leading to SAR-like patterns, where fewer parasite species are associated with relatively rare host species.

Given that, as we have argued above, SARs are expected to exist for parasites, estimates of the total number of unknown parasites will be problematic if the host species whose parasites or parasitoids have been investigated are disproportionately more common than a randomly selected species. This clearly seems to be the case. For example, an estimation of the global number of parasitoid species (Forbes et al. 2018) relied on four host genera, each coming from one of the four largest insect orders, and relied on data from thoroughly studied insect species. The data, however, also come from highly abundant animals. One genus (*Rhagoletis*: Diptera) is an important pest in agriculture and the three others (*Malacosoma*: Lepidoptera, *Dendroctonus*: Coleoptera and *Neodiprion*: Hymenoptera) are among the most important pests in forestry (Forbes et al. 2018). To take another example, Larsen et al. (2017) estimated the total number of parasites/commensal animals globally (with parasitoids not included in their definition of parasites) and argued that mites and nematodes are by far the most diverse groups. They estimated the number of parasitic/commensal mites (Larsen et al. 2017) based on several papers, but the two most recent studies (Knee et al. 2012, 2013) are based on bark beetles including the extremely abundant *Dendroctonus* beetles mentioned above. Larsen et al. (2017) also used several sources to generate nematode diversity estimates, but it is interesting that one of their highest estimated numbers of host-specific parasites again comes from a genus of bark beetles (*Ips*), which are important pests in forestry (Grucmanová and Holuša 2013). We note that major pests are not necessarily the most abundant species naturally, as for example seen in the Colorado potato beetle (*Leptinotarsa decemlineata*) which is a worldwide major pest today but apparently used to be a moderately common species with a spatially limited range (Casagrande 1985). All else being equal, however, it seems fair to assume that the abundances over evolutionarily relevant timescales of the insect species investigated for their parasite or parasitoid fauna are substantially larger than randomly selected insect species.

In summary, if there is a SAR in animal parasites and parasitoids (which seems theoretically very likely although it has yet to be convincingly shown empirically), the previous estimates of the number of parasites and parasitoids are likely to be vast overestimates given that they are based on some of the most abundant insect species.

4 | Estimates of the Number of Cryptic Species

A related problem is apparent in estimates of the number of cryptic species, that is, distinct species that cannot be distinguished based on their morphology (at least based on current knowledge). Several studies (e.g., Larsen et al. 2017; Li and Wiens 2023) have tried to identify, for a range of taxa, the average number of cryptic species per analysed morphological species in order to estimate the total number of undescribed cryptic species. A thorough study focused on cryptic insect species did acknowledge that their estimate could be an overestimate if cryptic species are more common in abundant and widespread morphological species and acknowledged that their data overrepresented such species (Li and Wiens 2023). They however went on to argue that their results would be unlikely to suffer from this bias (Li and Wiens 2023).

Yet, there are strong reasons to assume that species with available genetic data are disproportionally widespread, and that widespread species are disproportionally likely to contain cryptic species. Among mammals, a study found a clear pattern where morphological species suggested to comprise a number of cryptic species had on average a much larger range size than species without any such evidence (Parsons et al. 2022). This pattern may be even stronger than it appears in their data since species with genetic data (which is a requirement when looking for cryptic species) are on average more widespread than species without genetic data (Reddy 2014). We acknowledge that we cannot know if the relationship between range size and cryptic species is also present in arthropods, which are the most important group in estimates of global undescribed diversity, but there is no obvious reason why a relationship between range size and cryptic divergence should be vertebrate-specific.

There are also theoretical reasons why more widespread morphological species should be more likely to contain cryptic species. Currently accepted species generally correspond to clades representing one or more actual (but potentially cryptic) species. As long as there is not a perfect correlation between the rate of speciation and the rate of evolution of morphological characters that enable scientists to tell species apart, cryptic species should be more common in clades with higher speciation rates. A higher speciation rate for widespread species is supported by both theoretical models and empirical studies (e.g., Pigot et al. 2010; Smyčka et al. 2023), and widespread morphological species should therefore on average contain more cryptic species than geographically restricted ones.

In summary, we again have strong reasons to assume that the current estimates of the number of cryptic species are overestimates because they are based on morphological species which are disproportionately likely to contain cryptic species compared to a randomly selected morphologically defined species.

5 | Estimating the Effect Size

We have argued for the presence of a bias but have not yet tried to estimate the magnitude of it. The magnitude of the bias for parasites and for cryptic species may not yet be possible to estimate but for the case of plant-feeding insects some data are available, which will enable us to illustrate the potential magnitude of these biases. The first step is deciding on a mathematical function for the SAR. The two most frequently applied models to estimate the number of species (n_A in an area of size A) are the power model ($n_A = c_{\text{Pow}} \times A^{z_{\text{Pow}}}$) and the semi-log model ($n_A = z_{\text{Semilog}} \times \log(A) + c_{\text{Semilog}}$), where c and z are coefficients estimated from empirical data. Across true island datasets, the power law has typically been found to provide a marginally better fit than the semi-log model (Triantis et al. 2012). However, we will use the semi-log model here for the purpose of our approximation of the bias, as it allows for an easier transfer of the slope coefficient of the relationship (z_{Semilog}) across systems, as we show below. In this regard, we also note that the complex reality of diversity scaling can never be fully captured by SAR models with a limited number of parameters. Several important factors, such as the often-discussed small island effect whereby species richness initially increases at a slower rate for small islands (see e.g., Chiarucci et al. 2021), are necessarily omitted from these simple models and the results from them should therefore always be interpreted with caution.

If we only look at specialist species, it seems clear that there must be some lower bound where no associated species exist. More generalist species could in principle be associated with trees of any population size, but for most insects adding an additional host plant requires some adaptations, for example, associated with the plant's chemical defences or how adult insects identify host plants (Després et al. 2007). There must again be some lower bound where such adaptations are not evolutionarily advantageous to maintain. We acknowledge that this lower bound does not need to be the same for specialist and generalist species, and that a small number of hyper generalists exist (see e.g., Forister et al. 2014), which can be found on plants with any population size. For model simplicity we will however assume a single lower bound, which we indicate with k. Below this lower bound we will assume no associated insects.

Based on these definitions, we can predict the number of species in an area of size A as follows:

$$n_A = I \times \left(z_{\text{Semilog}} \times \log_{10}(A) + c_{\text{Semilog}}\right) \text{ where } \begin{cases} I = 1, \text{if } A > k \\ I = 0, \text{otherwise} \end{cases} \tag{1}$$

where k can either be specified based on empirical observation or set to the value above which the predicted species number is not negative, that is,

$$k = 10^{-c_{\text{Semilog}}/z_{\text{Semilog}}} \tag{2}$$

For plant-feeding insects, we can get potential values for the above equation based on the data from Brändle and Brandl (2002). After fitting a log-linear regression model to the Brändle and Brandl (2002) data and setting A to the abundance of a tree species, we obtain (multiple R-squared: 0.30, $p = 0.0046$) a slope $z_{\text{Semilog}} = 524$, an intercept $c_{\text{Semilog}} = -1489$ and threshold $k = 694$.

If, for the sake of argument, we assume that the species–area slope (z_{Semilog}) (with area here representing the abundance of a given tree species) is consistent across areas (although we acknowledge that it will vary across regions in reality), we can use this estimate and apply it to the Amazonian data, while still allowing for a different intercept (c_{Semilog}). Erwin (1982) estimated there to be 1200 insect species associated with *L. seemannii* and ter Steege et al. (2020) estimated the $\log_{10}$ population size of *L. seemannii* in the Amazon population to be 5.66 (0.45 million individuals). Using this and applying the semi-log model slope parameter calculated above (i.e., $z_{\text{Semilog}} = 524$), we can estimate the intercept (rounded to integer) $c_{\text{Semilog}} = -1764$, and the parameter $k = 2325$ (the minimum number of individuals that can sustain one insect species, here rounded to an integer), and can use these to predict the number of associated insects for tropical tree species using Equation (1).

ter Steege et al. (2020) used data from a large network of tree plots to estimate tree diversity in the Amazon. They found that the log transformed rank abundance curve was approximately linear and predicted a total diversity of about 15,000 tree species (between 13,887 and 17,020) of which the majority was unsampled, that is, did not occur in any of their plots. Using a linear model to predict the abundances of unsampled species, we find that 5758 tree species (almost 40%) have predicted abundance lower than our estimated threshold k. Applying Equation (1) to the predicted tree abundances we further obtain a mean insect diversity per tree species of 798. This suggests that taking the estimate from *L. seemannii* (i.e., 1200) to extrapolate insect diversity without accounting for the SAR would lead to a 50% overestimation.

The transference of SAR slopes between temperate systems (i.e., $z_{\text{Semilog}} = 524$) and tropical ones is of course problematic and only intended as an illustrative example of the magnitude of the problem. Yet, we find that repeating our predictions of mean insect diversity per tree species under a range of slopes, spanning from 524/2 to 524 × 2, still result in estimates substantially lower than the 1200 species measured for *L. seemannii* (Figure 2). Under this range of scenarios, assuming 1200 insect species across all tree species would lead to overestimations ranging from 25% to 52%.

We can also look at an empirical distribution of SAR slopes from an analysis of 601 true island datasets covering multiple taxa (Triantis et al. 2012). These slopes are derived from the analysis of biogeographical systems which may respond differently from the evolutionary relationship between trees and herbivorous insects. We also note that there may be sampling biases associated with these, for example, as a function of relatively more intense sampling on smaller oceanic islands. Yet they may still represent a useful comparison to assess if the SAR slope (of the semi-log model) we estimated for temperate trees is comparable to those from biogeographical datasets. However, initial examination of the slopes revealed that the slope is clearly taxon-specific. This is most obvious when thinking about taxonomically nested datasets where the slopes necessarily increase for higher taxonomic groups. Taking the simplest theoretical example, if the diversity of two groups G and H are given by $n_{\text{G,A}} = z_{\text{G}} \times \log(A) + c_{\text{G}}$ and $n_{\text{H,A}} = \text{z}_{\text{H}} \times \log(A) + c_{\text{H}}$, then the combined diversity is given by $n_{\text{G,A}} + n_{\text{H,A}} = (z_{\text{G}} + z_{\text{H}}) \times \log(A) + c_{\text{G}} + c_{\text{H}}$. Thus, semi-log model slopes are not comparable among organism groups that vary in the number of species they contain. Yet, if similar underlying biological processes were operating, the ratio between the maximum predicted diversity (e.g., the number of species associated with the largest area or most abundant host species, here indicated with n_{Amax}) and the slope may be approximately constant. Using the 601 datasets and Equation (1), we calculated the ratio between the slope z_{Semilog} and the observed maximum diversity value n_{Amax} and find that indeed this relationship is unimodal and relatively constrained, with a median of 4.39% and 75% of the values between 2.3 and 7.6 (Figure 3). Going back to the insect example in Amazonia and using the estimated $n_{\text{Amax}} = 3303$ for the most abundant species in the dataset (*Eschweilera coriacea*) and $z_{\text{Semilog}} = 524$, we find that the ratio for this dataset was 6.3, which is well within the range of empirical measurements (Figure 3). This suggests that the slope of ecological species–area relationships is relatively well constrained and that our estimates of Amazonian insect diversity per tree species fall within a plausible range, despite the extrapolations made from temperate data.

These estimates are based on the semi-log model where we could use empirical values, but there are several alternative models for the SAR (Triantis et al. 2012), with the power model being the most obvious alternative candidate model. However, we cannot

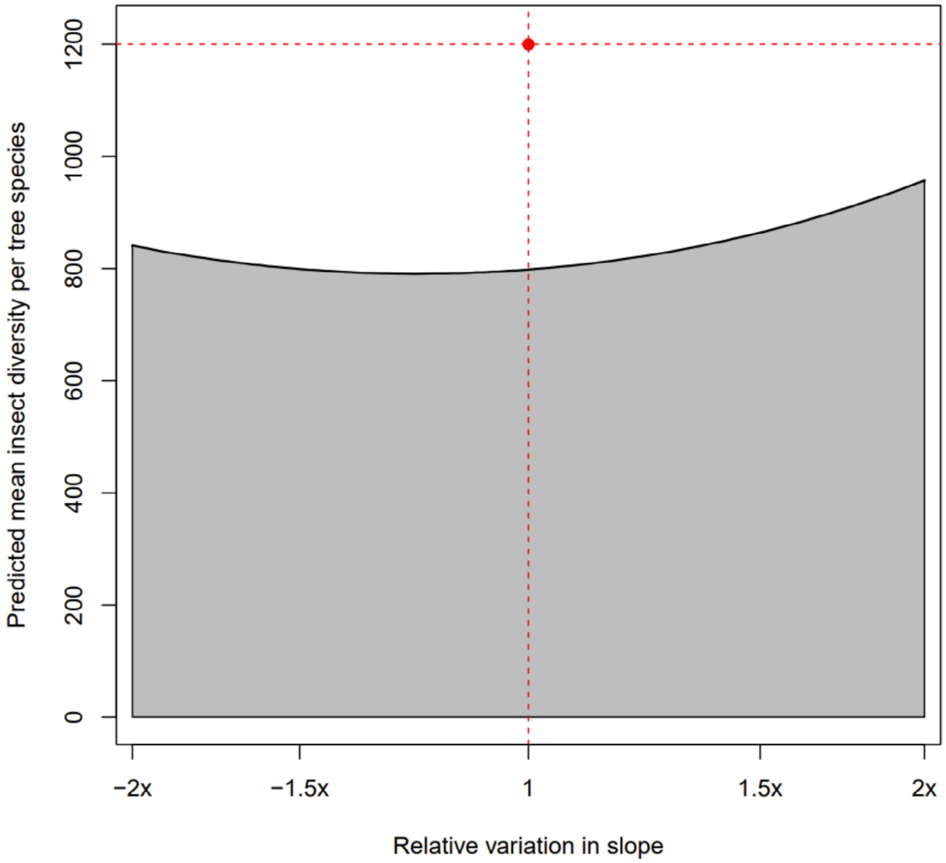

FIGURE 2 | Estimated average insect diversity per tree species under a range of SAR slope (z_{Semilog} parameter of the semi-logSAR model) values spanning from 524/2 to 524 × 2. These averages are estimated while accounting for the SAR using Equation (1) across all Amazonian tree species. The horizontal dashed line shows the 1200 insect species associated with the tree species *Luehea seemannii* and clearly indicates that using this value as a proxy for mean insect diversity per tree species would lead to a substantially higher estimate of total diversity, across a wide range of slopes, here $z_{\mathrm{Semilog}} \in \{262, 1048\}$. The overestimation using the slope of the temperate tree SARs, $z_{\mathrm{Semilog}} = 524$, is 50% (red circle).

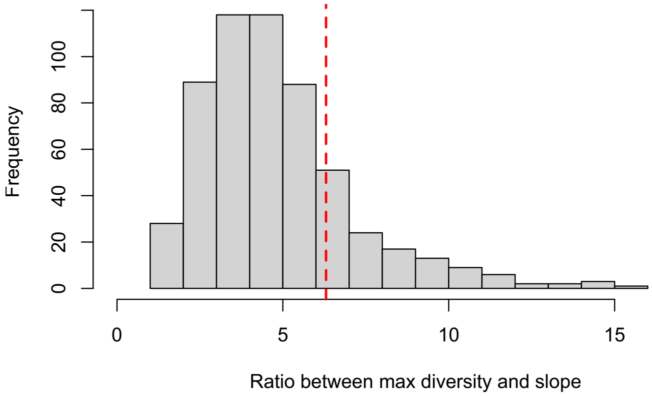

FIGURE 3 | Frequency distribution of the empirical ratio between observed maximum species richness and the slope of the semi-log SAR model. The data come from 601 true island datasets from Triantis et al. (2012), where each dataset represents the number of species of a given taxon on islands in an archipelago. The y-axis represents the frequency among the 601 studies considered here. For plotting purposes, we included datasets within the 2.5%–97.5% quantiles, resulting in the exclusion of 32 points, that is, 5% (the full range can be seen in Figure S1). The estimated ratio for insect diversity on Amazonian trees (dashed red line), although based on extrapolations from temperate trees, is clearly within the range of empirical island SARs.

reliably transfer the exponent (z_{Pow}) of the power model from temperate to tropical systems. If we fit a power model to the German study discussed above (Brändle and Brandl 2002), we get an exponent of 1. If we tried transferring this exponent directly, we would get the nonsensical result of 35.4 million insect species per tree species. We can instead try a suggested value of the exponent from the literature, such as 0.2, which has been suggested as plausible for continental settings in conventional SARs (Kunin et al. 2018). Using this exponent, we get a mean insect diversity per tree species of 1095, which is just 9% lower than Erwin's (1982) estimate of 1200 insects per tree species. It has recently been suggested that the true SAR in many cases is intermediate between the semi-log and the power model (Carey et al. 2023). Taking all of this into account, the best available data suggest that the extrapolated diversity estimates may be up to 50% too high due to the failure to integrate species–area relationships into the estimates, but there is considerable uncertainty to consider.

We also note that all the above estimates of the bias are conditional on the population size of *L. seemannii*, which is known to be an underestimate. For one thing, our population estimate only relates to the Amazonian abundance, whereas the majority of the distribution of the species is in Central America (Condit 2019). An analysis has suggested that about ¼ of all tree species in the Amazon have the majority of their range outside the Amazon (Feeley 2016). The effect of this is that not integrating SAR dynamics into diversity estimates is potentially an even bigger problem than the numbers and analyses above suggest, although we do not have adequate data to estimate the extent to which this is the case.

In conclusion, it is important to highlight that there are several reasons why Erwin's (1982) predictions are overestimates and we are not arguing that the SAR related issue presented here is the most important or only source of bias. For example, *L. seemannii* could be atypical for other reasons (e.g., in terms of its phytochemistry), and the fraction of species that were considered monophagous was based on a rough and likely inaccurate rule of thumb. Indeed, the predicted number of phytofagous insects has been drastically lowered following Erwin's (1982) estimate of 30 million. Later analyses estimated the number to be one order of magnitude lower, with a reported 90% confidence interval between 2 and 7 million (Hamilton et al. 2010). However, the identification and discussion of other potentially larger sources of bias does not negate the importance of highlighting the potential SAR related bias, particularly as this likely also affects other elements of global richness estimation (including but not limited to estimates of total parasite diversity and estimates of cryptic species numbers).

6 | Ways Forward

Here, we have addressed three cases in which ignoring the SAR results in biased global diversity estimates, but it should be noted that they may interact with and potentially strengthen one another. As an example, we consider the estimate of the total number of animal species generated by Larsen et al. (2017). They started out with an estimate of 6.8 million insect species from Stork et al. (2015). This estimate is not directly related to the SAR, but one of the arguments for its credibility used by Stork et al. (2015) was its similarity with estimates based on plant-associated insects, which, as we argued above, are near guaranteed to be overestimates. Larsen et al. (2017) multiplied this estimate of morphological species diversity with an estimate of the number of cryptic species per insect, which we also discussed above is near guaranteed to be an overestimate. This was then multiplied with an estimate of the number of mites per insect, which suffers from the same overestimation bias. Finally, the total number of arthropods (based on insects and mites) was multiplied by the average number of nematodes per arthropod, which we contend will be biased in the same direction. Although the undescribed diversity of these groups is unquestionably vast, extrapolation biases potentially lead to a cascade of error propagation and amplification, resulting in unrealistic biodiversity estimates.

The biases we focus on are restricted to classical approaches, leading to the question of whether the easiest solution to the problem is using eDNA to estimate global biodiversity instead. We, however, consider this unlikely. As DNA sequencing is rapidly becoming cheaper and easier, some scientists have argued that barcoding approaches may resolve uncertainties in the number of species (Hebert et al. 2016). DNA barcoding may prevent some of the issues leading to error amplification just mentioned, but it alone will not lead to correctly estimated total species diversity for similar reasons to those we have addressed here. Many species are extremely narrowly distributed (and/or very rare), and it seems unrealistic to assume that all or even a large number of species can be picked up by generalized sampling procedures. The distribution of range-restricted species is at least somewhat predictable (Sandel et al. 2011), and this could theoretically be incorporated into a sampling regime, but each area only contains a moderately small number of narrow-ranged species (Jetz and Rahbek 2002). It seems likely that eDNA sequencing can solve many issues but probably not the ones we address here, at least not unless an unrealistically dense global sampling regime is undertaken, involving sampling every few kilometres apart across the globe.

There is a rich literature on upscaling biodiversity, with many proposed approaches (see e.g., Kunin et al. 2018), and we are not arguing that all species richness extrapolation approaches are flawed. Rather, our analyses show that many estimates of global diversity—particularly those based on extrapolating based on a few well studies species—may be substantially biased. In our opinion, we cannot get reliable estimates of global diversity based on such generalizing without integrating the species–area relationship into the analyses, and to do so we may require large amounts of field work. The plant-eating insects we discussed first may be the pattern where empirical data are easiest to obtain. Different research groups have already undertaken each half of this work for the insect-plant relationship. This was unfortunately done on separate tropical continents—as we noted above, ter Steege et al. (2020) estimated population sizes of all tree species in the Amazon and in another study Novotny et al. (2002) identified all plant associated insects on 51 different tree species in New Guinea—but it shows that the data needed to analytically solve this can be obtained and we hope that this paper may motivate future empirical work trying to do so. Although with the available data it remains difficult to provide robust estimates of the magnitude of the problems shown here, we think existing evidence points to substantial overestimations of current global diversity estimates.

Author Contributions

S.F. initiated the study and wrote the first draft of the paper. D.S. and S.F. analysed the data. All authors contributed to the design of the study and to the writing.

Acknowledgements

We thank Dr. Kostas A. Triantis (University of Athens) for providing us with unpublished data on the distribution of SAR slopes.

Funding

S.F. thanks the Swedish Research Council Vetenskapsrådet (no. 2021-04690) for support. D.S. acknowledges ETH Zurich for funding and the Swedish Foundation for Strategic Environmental Research MISTRA within the framework of the research programme BIOPATH (F 2022/1448). TM acknowledges the University of Birmingham's GEES Travel Fund for support during the preparation of this manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All code and data are uploaded to Dryad http://datadryad.org/share/8onHc0lst9fH63XdIJPHCvCqyxP35hMzh9wHMlGyXIU.

References

Baldridge, E., D. J. Harris, X. Xiao, and E. P. White. 2016. "An Extensive Comparison of Species-Abundance Distribution Models." *PeerJ* 4: e2823. https://doi.org/10.7717/peerj.2823.

Basset, Y., G. A. Samuelsen, A. Allison, and S. E. Miller. 1996. "How Many Species of Host-Specific Insects Feed on a Species of Tropical Tree?" *Biological Journal of the Linnean Society* 59: 201–216. https://doi.org/10.1111/j.1095-8312.1996.tb01461.x.

Bebber, D. P., F. H. C. Marrriott, K. J. Gaston, S. A. Harris, and R. W. Scotland. 2007. "Predicting Unknown Species Numbers Using Discovery Curves." *Proceedings of the Royal Society B* 274: 1651–1658. https://doi.org/10.1098/rspb.2007.0464.

Black, F. L. 1966. "Measles Endemicity in Insular Populations: Critical Community Size and Its Evolutionary Implication." *Journal of Theoretical Biology* 11: 207–211. https://doi.org/10.1016/0022-5193(66)90161-5.

Brändle, M., and R. Brandl. 2002. "Species Richness of Insects and Mites on Trees: Expanding Southwood." *Journal of Animal Ecology* 70: 491–504. https://doi.org/10.1046/j.1365-2656.2001.00506.x.

Callaghan, C. T., L. Borda-de-Água, R. van Klink, R. Rozzi, and H. M. Pereira. 2023. "Unveiling Global Species Abundance Distributions." *Nature Ecology & Evolution* 7: 1600–1609. https://doi.org/10.1038/s41559-023-02173-y.

Callaghan, C. T., S. Nakagawa, and W. K. Cornwell. 2021. "Global Abundance Estimates for 9,700 Bird Species." *Proceedings of the National Academy of Sciences of the United States of America* 118: e2023170118. https://doi.org/10.1073/pnas.2023170118.

Carey, M., J. Boland, and G. Keppel. 2023. "Generalized Logarithmic Species–Area Relationship Resolves the Arrhenius-Gleason Debate." *Environmental Modeling & Assessment* 28: 491–499. https://doi.org/10.1007/s10666-023-09873-6.

Caribherp. 2025. https://www.caribherp.org/.

Casagrande, R. A. 1985. "The 'Iowa' Potato Beetle, Its Discovery and Spread to Potatoes." *Bulletin of the Entomological Society of America* 31: 27–29. https://doi.org/10.1093/besa/31.2.27.

Chiarucci, A., R. Guarino, S. Pasta, et al. 2021. "Species–Area Relationship and Small-Island Effect of Vascular Plant Diversity in a Young Volcanic Archipelago." *Journal of Biogeography* 48: 2919–2931. https://doi.org/10.1111/jbi.14253.

Condit, R. 2019. "*Luehea seemannii*." The IUCN Red List of Threatened Species 2019: e.T151972680A151972697.

Connell, J. H. 1971. "On the Role of Natural Enemies in Preventing Competitive Exclusion in Some Marine Animals and in Rain Forest Trees." In *Dynamics of Populations*, edited by P. J. den Boer and G. R. Gradwell, 298–310. Centre for Agricultural Publications and Documentation.

Costello, M. J., R. M. May, and N. E. Stork. 2013. "Can We Name Earth's Species Before They Go Extinct?" *Science* 339: 413–416. https://doi.org/10.1126/science.1230318.

Després, L., J. P. David, and C. Gallet. 2007. "The Evolutionary Ecology of Insect Resistance to Plant Chemicals." *Trends in Ecology & Evolution* 22: 298–307. https://doi.org/10.1016/j.tree.2007.02.010.

Enquist, B. J., X. Feng, B. Boyle, et al. 2019. "The Commonness of Rarity: Global and Future Distribution of Rarity Across Land Plants." *Science Advances* 5: eaaz0414. https://doi.org/10.1126/sciadv.aaz0414.

Erwin, T. L. 1982. "Tropical Forests: Their Richness in Coleoptera and Other Arthropod Species." *Coleopterists Bulletin* 36: 74–75.

Feeley, K. J. 2016. "Commentary: Estimating the Global Conservation Status of More Than 15,000 Amazonian Tree Species." *Frontiers of Ecology and Evolution* 4: 59. https://doi.org/10.3389/fevo.2016.00059.

Forbes, A. A., R. K. Bagley, M. A. Beer, A. C. Hippee, and H. A. Widmayer. 2018. "Quantifying the Unquantifiable: Why Hymenoptera, Not Coleoptera, Is the Most Speciose Animal Order." *BMC Ecology* 18: 21. https://doi.org/10.1186/s12898-018-0176-x.

Forister, M. L., V. Novotny, A. K. Panorska, et al. 2014. "The Global Distribution of Diet Breadth in Insect Herbivores." *Proceedings of the National Academy of Sciences of the United States of America* 112: 442–447. https://doi.org/10.1073/pnas.1423042112.

Gaston, K. J. 1998. "Species-Range Size Distributions: Products of Speciation, Extinction and Transformation." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 353: 219–230. https://doi.org/10.1098/rstb.1998.0204.

Griffin, B. D., and J. M. Drake. 2008. "Effects of Habitat Quality and Size on Extinction in Experimental Populations." *Proceedings of the Royal Society of London. Series B: Biological Sciences* 275: 2251–2256. https://doi.org/10.1098/rspb.2008.0518.

Grucmanová, S., and J. Holuša. 2013. "Nematodes Associated With Bark Beetles, With Focus on the Genus Ips (Coleoptera: Scolytinae) in Central Europe." *Acta Zoologica Bulgarica* 65: 547–556.

Hamilton, A. J., Y. Basset, K. K. Benke, et al. 2010. "Quantifying Uncertainty in Estimation of Tropical Arthropod Species Richness." *American Naturalist* 176: 90–95. https://doi.org/10.1086/652998.

Hawksworth, D. L. 1991. "The Fungal Dimension of Biodiversity: Magnitude, Significance, and Conservation." *Mycological Research* 95: 641–655. https://doi.org/10.1016/S0953-7562(09)80810-1.

Hebert, P. D., S. Ratnasingham, E. V. Zakharov, et al. 2016. "Counting Animal Species With DNA Barcodes: Canadian Insects." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 371: 20150333. https://doi.org/10.1098/rstb.2015.0333.

Janzen, D. H. 1968. "Host Plants as Islands in Evolutionary and Contemporary Time." *American Naturalist* 102: 592–595. https://doi.org/10.1086/282574.

Janzen, D. H. 1970. "Herbivores and the Number of Tree Species in Tropical Forests." *American Naturalist* 104: 501–528. https://doi.org/10.1086/282687.

Jetz, W., and C. Rahbek. 2002. "Geographic Range Size and Determinants of Avian Species Richness." *Science* 297: 1548–1551. https://doi.org/10.1126/science.1072779.

Kamiya, T., K. O'Dwyer, S. Nakagawa, and R. Poulin. 2014. "What Determines Species Richness of Parasitic Organisms? A Meta-Analysis Across Animal, Plant and Fungal Hosts." *Biological Reviews* 89: 123–134. https://doi.org/10.1111/brv.12046.

Knee, W., F. Beaulieu, J. H. Skevington, S. Kelso, A. I. Cognato, and M. R. Forbes. 2012. "Species Boundaries and Host Range of Tortoise Mites (Uropodoidea) Phoretic on Bark Beetles (Scolytinae), Using Morphometric and Molecular Markers." *PLoS One* 7: e47243. https://doi.org/10.1371/journal.pone.0047243.

Knee, W., M. R. Forbes, and F. Beaulieu. 2013. "Diversity and Host Use of Mites (Acari: Mesostigmata, Oribatida) Phoretic on Bark Beetles (Coleoptera: Scolytinae): Global Generalists, Local Specialists?" *Annals of the Entomological Society of America* 106: 339–350. https://doi.org/10.1603/AN12092.

Kunin, W. E., J. Harte, F. He, et al. 2018. "Upscaling Biodiversity: Estimating the Species–Area Relationship From Small Samples." *Ecological Monographs* 88: 170–187. https://doi.org/10.1002/ecm.1284.

Larsen, B. B., E. C. Miller, M. K. Rhodes, and J. J. Wiens. 2017. "Inordinate Fondness Multiplied and Redistributed: The Number of Species on Earth and the New Pie of Life." *Quarterly Review of Biology* 92: 230–265. https://doi.org/10.1086/693564.

Li, X., and J. J. Wiens. 2023. "Estimating Global Biodiversity: The Role of Cryptic Insect Species." *Systematic Biology* 72: 391–403. https://doi.org/10.1093/sysbio/syac069.

Lomolino, M. V. 1990. "The Target Area Hypothesis: The Influence of Island Area on Immigration Rates of Non-Volant Mammals." *Oikos* 57: 297–300. https://doi.org/10.2307/3565957.

MacArthur, R., and E. O. Wilson. 1967. *The Theory of Island Biogeography*. Princeton University Press.

May, R. M. 1988. "How Many Species Are There on Earth?" *Science* 241: 1441–1449. https://doi.org/10.1126/science.241.4872.1441.

Mora, C., D. P. Tittensor, S. Adl, A. G. B. Simpson, and B. Worm. 2011. "How Many Species Are There on Earth and in the Ocean?" *PLoS Biology* 9: e1001127. https://doi.org/10.1371/journal.pbio.1001127.

Niskanen, T., R. Lücking, A. Dahlberg, et al. 2023. "Pushing the Frontiers of Biodiversity Research: Unveiling the Global Diversity, Distribution, and Conservation of Fungi." *Annual Review of Environment and Resources* 48: 149–176. https://doi.org/10.1146/annurev-environ-112621-090937.

Novotny, V., Y. Basset, S. E. Miller, et al. 2002. "Low Host Specificity of Herbivorous Insects in a Tropical Forest." *Nature* 416: 841–844. https://doi.org/10.1038/416841a.

Parsons, D. J., T. A. Pelletier, J. G. Wieringa, and B. C. Carstens. 2022. "Analysis of Biodiversity Data Suggests That Mammal Species Are Hidden in Predictable Places." *Proceedings of the National Academy of Sciences of the United States of America* 119: e2103400119. https://doi.org/10.1073/pnas.2103400119.

Payne, J. L., and S. Finnegan. 2007. "The Effect of Geographic Range on Extinction Risk During Background and Mass Extinction." *Proceedings of the National Academy of Sciences of the United States of America* 104: 10506–10511. https://doi.org/10.1073/pnas.0701257104.

Pigot, A. L., A. B. Phillimore, I. P. F. Owens, and C. D. L. Orme. 2010. "The Shape and Temporal Dynamics of Phylogenetic Trees Arising From Geographic Speciation." *Systematic Biology* 59: 660–673. https://doi.org/10.1093/sysbio/syq058.

Plants of the World Online (POWO). 2024. "Facilitated by the Royal Botanic Gardens, Kew." Published on the Internet. http://www.plantsoftheworldonline.org/.

Poulin, R., and S. Morand. 2000. "The Diversity of Parasites." *Quarterly Review of Biology* 75: 277–292. https://doi.org/10.1086/393500.

Reddy, S. 2014. "What's Missing From Avian Global Diversification Analyses?" *Molecular Phylogenetics and Evolution* 77: 159–166. https://doi.org/10.1016/j.ympev.2014.04.023.

Sandel, B., L. Arge, B. Dalsgaard, et al. 2011. "The Influence of Late Quaternary Climate-Change Velocity on Species Endemism." *Science* 334: 660–664. https://doi.org/10.1126/science.1210173.

Scheffers, B. R., L. N. Joppa, S. L. Pimm, and W. F. Laurance. 2012. "What We Know and Don't Know About Earth's Missing Biodiversity." *Trends in Ecology & Evolution* 27: 501–510. https://doi.org/10.1016/j.tree.2012.05.008.

Smyčka, J., A. Toszogyova, and D. Storch. 2023. "The Relationship Between Geographic Range Size and Rates of Species Diversification." *Nature Communications* 14: 5559. https://doi.org/10.1038/s41467-023-41225-6.

Southwood, T. R. E. 1960. "The Abundance of the Hawaiian Trees and the Number of Their Associated Insect Species." *Proceedings of the Hawaiian Entomological Society* 17: 299–303.

Southwood, T. R. E. 1961. "The Number of Species of Insect Associated With Various Trees." *Journal of Animal Ecology* 30: 1–8.

Stork, N. E., J. McBroom, C. Gely, and A. J. Hamilton. 2015. "New Approaches Narrow Global Species Estimates for Beetles, Insects, and Terrestrial Arthropods." *Proceedings of the National Academy of Sciences of the United States of America* 112: 7519–7523. https://doi.org/10.1073/pnas.1502408112.

Strona, G., and S. Fattorini. 2014. "A Few Good Reasons Why Species-Area Relationships Do not Work for Parasites." *BioMed Research International* 2014: 271680. https://doi.org/10.1155/2014/271680.

ter Steege, H., P. I. Prado, R. A. F. de Lima, et al. 2020. "Biased-Corrected Richness Estimates for the Amazonian Tree Flora." *Scientific Reports* 10: 10130. https://doi.org/10.1038/s41598-020-66686-3.

Triantis, K. A., F. Guilhaumon, and R. J. Whittaker. 2012. "The Island Species–Area Relationship: Biology and Statistics." *Journal of Biogeography* 39: 215–231. https://doi.org/10.1111/j.1365-2699.2011.02652.x.

Wiens, J. J. 2023. "How Many Species Are There on Earth? Progress and Problems." *PLoS Biology* 21: e3002388. https://doi.org/10.1371/journal.pbio.3002388.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** Empirical ratio between maximum species richness and slopes of logarithmic SAR models. Data come from 601 true island datasets from Triantis et al. (2012), where each dataset represents the number of species of a given taxon on islands in an archipelago. The value estimated based on temperate trees is shown as a vertical red line, and the median value across all SARs is shown as a vertical blue line.