Volume 92, No. 3 September 2017

The Quarterly Review of Biology

INORDINATE FONDNESS MULTIPLIED AND REDISTRIBUTED: THE NUMBER OF SPECIES ON EARTH AND THE NEW PIE OF LIFE

Brendan B. Larsen*
Department of Ecology and Evolutionary Biology, University of Arizona
Tucson, Arizona 85721-0081 USA
eail: bblarsen@email.arizona.edu

Elizabeth C. Miller
Department of Ecology and Evolutionary Biology, University of Arizona
Tucson, Arizona 85721-0081 USA
e-mail: ecmiller@email.arizona.edu

Matthew K. Rhodes
Department of Ecology and Evolutionary Biology, University of Arizona
Tucson, Arizona 85721-0081 USA
e-mail: mkrhodes@email.arizona.edu

John J. Wiens†
Department of Ecology and Evolutionary Biology, University of Arizona
Tucson, Arizona 85721-0081 USA
e-mail: wiensj@email.arizona.edu

keywords
arthropods, bacteria, biodiversity, cryptic species, parasites, species richness

* Authorship order is alphabetical
† Corresponding author

The Quarterly Review of Biology, September 2017, Vol. 92, No. 3

0033-5770/2017/9203-0001$15.00

ABSTRACT

The number of species on Earth is one of the most fundamental numbers in science, but one that remains highly uncertain. Clearly, more species exist than the present number of formally described species (approximately 1.5 million), but projected species numbers differ dramatically among studies. Recent estimates range from about 2 million species to approximately 1 trillion, but most project around 11 million species or fewer. Numerous studies have focused on insects as a major component of overall richness, and many have excluded other groups, especially non-eukaryotes. Here, we re-estimate global biodiversity. We also estimate the relative richness of the major clades of living organisms, summarized as a "Pie of Life." Unlike many previous estimates, we incorporate morphologically cryptic arthropod species from molecular-based species delimitation. We also include numerous groups of organisms that have not been simultaneously included in previous estimates, especially those often associated with particular insect host species (including mites, nematodes, apicomplexan protists, microsporidian fungi, and bacteria). Our estimates suggest that there are likely to be at least 1 to 6 billion species on Earth. Furthermore, in contrast to previous estimates, the new Pie of Life is dominated by bacteria (approximately 70–90% of species) and insects are only one of many hyperdiverse groups.

INTRODUCTION

HOW many species are there on Earth? This is a fundamental question in science, but one that remains far from resolved. It is widely agreed that the number of described species (approximately 1.5 million species; Roskov et al. 2014) underestimates actual global richness, but the extent of this underestimation remains unclear. Projections of global biodiversity have ranged from as low as ~2 million species (Costello et al. 2012), up to ~100 million (e.g., Ehrlich and Wilson 1991; May 1992; Lambshead 1993), or even ~1 trillion (Locey and Lennon 2016).

In some ways, there is an encouraging trend in that some estimates appear to have stabilized. For example, some larger estimates of global richness have centered on tropical insects (e.g., approximately 30 million; Erwin 1982). Several recent studies have now converged at ~6 million insect species (Novotny et al. 2002; Basset et al. 2012; Stork et al. 2015), and ~11 million or fewer species overall (Mora et al. 2011; Costello et al. 2012). However, most estimates do not incorporate morphologically cryptic species discovered through molecular analyses, which might dramatically increase richness (e.g., Bickford et al. 2007; Adams et al. 2014). Many papers also exclude major branches of the Tree of Life, such as bacteria. Further, some studies have included bacteria but have estimated their richness to be very low (e.g., Mora et al. 2011). Other studies have focused specifically on estimating microbial species richness, but have not simultaneously addressed the richness of other groups. Some studies have estimated the presence of 10^7 to 10^9 bacterial species globally (Dykhuizen 1998; Curtis et al. 2002), whereas others have suggested that such estimates are too high (Schloss and Handelsman 2004). An exciting recent study used scaling laws to project the presence of ~1 trillion microbial species (Locey and Lennon 2016), but did not address to what groups these microbes belong (e.g., bacteria versus archaeans versus microbial fungi versus protists), where they occur (e.g., soil versus endosymbiotic), or the richness of other groups.

A topic that is closely related to the number of species on Earth is the relative species richness of different clades, sometimes presented in a pie diagram (e.g., Wilson 1992; Figure 1). We refer to such a diagram as a "Pie of Life" (parallel to the Tree of Life). Relatively few studies have presented explicit estimates of the Pie of Life relative to the many that have estimated global species richness. Interestingly, some estimates of the relative richness of clades have remained similar to those based on described species (e.g., Wilson 1992), even with dramatic increases in projected richness overall (e.g., Mora et al. 2011; Figure 1).

Here, we provide new estimates of global species richness and the Pie of Life. Our approach differs from most previous studies in three main ways. First, as much as possible, we include all major groups of living organisms, including bacteria (but not viruses, which are not necessarily alive). Second, we

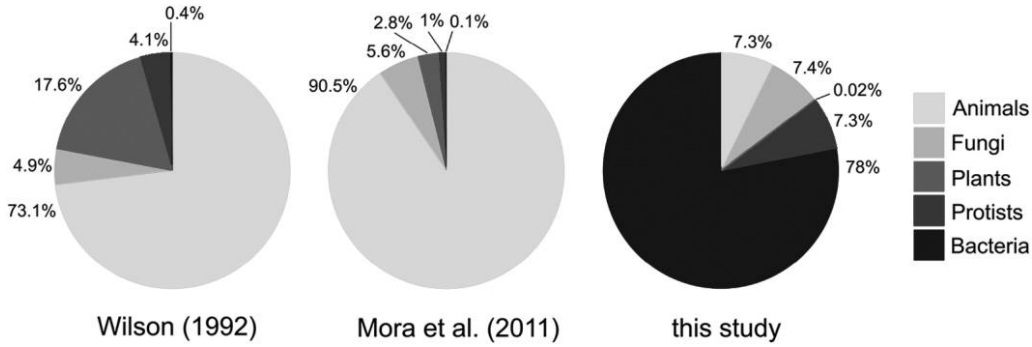

Figure 1. Traditional and New Estimates of the Pie of Life

The pie on the left shows a traditional estimate of the relative richness of different groups of organisms based on numbers of described species (Wilson 1992), the middle shows an estimate based on projected richness of different groups (Mora et al. 2011), and the pie on the right shows estimates based on the projected richness of different groups in the present study. The pie on the right is estimated based on Scenario 1 (Table 1), but other scenarios and assumptions give very similar estimates of the relative richness of different groups (Tables 1–4), even as total species richness changes. See the online edition for a color version of this figure.

incorporate morphologically cryptic species as part of our estimation, as revealed by molecular data from multiple genetic loci and the use of rigorous, well-established, coalescent-based species delimitation methods. Third, we emphasize groups that are associated with insects (e.g., as parasites or endosymbiotes), especially those groups with species that are relatively specific to each insect host species. These groups may have the strongest influence on overall species numbers. Estimates of insect richness based on morphologically defined species appear to have stabilized (approximately 6 million; Novotny et al. 2002; Basset et al. 2012; Stork et al. 2015), providing an important anchor for other projections. Our estimates suggest that there may be more than a billion species on Earth, that global richness is dominated by bacteria, and that the Pie of Life is very different from traditional estimates.

Methods

estimating animal richness

To estimate overall animal richness, we first used the mean estimated number of arthropod species from Stork et al. (2015), which is ~6.8 million (consisting mostly of insects). This number does not explicitly include mites that are associated with insects.

Next, we estimated the mean number of cryptic species for each morphologically defined arthropod species (for details see Appendix S1; Tables S1 and S2). This estimate is based on a systematic search for studies that applied a rigorous, well-established method of species delimitation to multilocus DNA sequence data, focusing on species that were sampled from multiple populations across their geographic ranges. We conducted a systematic search of the literature for studies from 2013–2015 that used the Bayesian Phylogenetics and Phylogeography method (BPP; Yang and Rannala 2010) to delimit cryptic species in arthropods (details in Appendix S1). The BPP method is standard and considered to be relatively accurate (e.g., Camargo et al. 2012; Carstens et al. 2013). For each of 16 studies, we counted the number of unnamed species supported by BPP relative to the number of described species included in the study. Note that almost all cryptic species occupied only parts of the geographic range of a described species. Thus, assessing cryptic diversity based on a single location or geographic region would miss most of these cryptic species.

We then multiplied the mean number of cryptic species per described species per study (5.9, rounded up to 6) by the projected number of arthropod species based on mor-

phological criteria (approximately 6.8 million). This yielded our baseline number of arthropod species (40.8 million). However, this estimate of six cryptic species might be an overestimate (see below and Appendix S1 for further discussion). First, an alternative calculation is to divide the total number of cryptic species by the total number of described species across all studies (rather than study-specific means), yielding two cryptic species per named species. Second, phylogeographic studies may focus preferentially on wide-ranging taxa with many cryptic species, such that our pool of species is not representative of all arthropod species. Third, some authors have suggested that BPP may overestimate species diversity if there is a time lag between genetic fragmentation and completion of speciation (Sukumaran and Knowles 2017). However, fragmentation may be the primary cause of allopatric speciation (Wiens 2004) and the evolution of intrinsic reproductive isolation may be decoupled from the origin of new species (e.g., Wiens et al. 2006; Rabosky and Matute 2013). Overall, given these considerations, we also explored smaller ratios of cryptic to morphologically distinct species, specifically, two and zero cryptic species (see below). Again, we emphasize that our results include the possibility that there are no cryptic arthropod species at all.

Next, we assumed that each of these arthropod species has (on average) at least one associated mite species (for justification see Appendix S2). We also explored a smaller ratio of arthropods to mite species (see below, Scenario 4). Assuming a 1:1 ratio of mites to other arthropods yields a total of 81.6 million arthropod species.

We then assumed an average of one nematode species per arthropod species (for justification see Appendix S2), yielding a total of 81.6 million nematode species. We added these estimates for arthropods and nematodes for a total of 163.2 million animal species. Of course, arthropods and nematodes are not the only animals, but recent projections suggest that it is unlikely that all other animal groups combined would have much more than a million species (e.g., Scheffers et al. 2012; see Appendix S2). Even a substantially larger estimate (e.g., 10 times this) would have relatively little impact on overall animal richness, given the projected numbers of arthropods and nematodes.

PLANTS

There are presently ~300,000 described plant species (Roskov et al. 2014). Quantitative estimates suggest that the actual number is not substantially higher than this (e.g., approximately 320,000 species; Mora et al. 2011; about 330,000–360,000 species; Joppa et al. 2011). We used an intermediate estimate of ~340,000 species. Importantly, even estimates that were two or three times this value would have little impact overall, given how small this number is relative to those for other groups.

ESTIMATING FUNGAL RICHNESS

For fungi, we assumed that each animal species contains (on average) one microsporidian species, yielding 163.2 million species (for justification see Appendix S3). To this, we added the estimated number of soil fungi (2.4 million; Appendix S3) to yield 165.6 million fungal species. We recognize that other fungal groups might also have very high species richness (e.g., endophytes, entomopathogenic fungi), but their host specificity is not sufficiently well established for us to estimate and include their richness (Appendix S3). Overall, we conclude that fungal richness is much higher than previously estimated, and so these additional sources of fungal richness would clearly not overturn that conclusion, but would only reinforce it further.

ESTIMATING PROTIST RICHNESS

For protists, we assumed that each animal species hosts (on average) at least one apicomplexan species (for justification see Appendix S4), yielding 163.2 million apicomplexans. Free-living protist richness might also be substantial, but possibly not much greater than 1 million species (Appendix S4). Therefore, we did not include this number in our estimates. Nevertheless, much higher free-living protist richness would only further support our conclusions that protist richness (and

overall richness) is much higher than presently estimated.

ARCHAEANS

We did not include archaeans as a separate clade in our overall estimates of global richness. Archaeans currently include less than 1000 described species (Roskov et al. 2014), and we are not aware of specific estimates suggesting that their undescribed diversity is many orders of magnitude higher. Specifically, given our overall results, there would need to be hundreds of millions of undescribed archaean species to strongly impact our global richness estimates (or substantially alter the Pie of Life). Archaea can be very abundant be in the ocean (Karner et al. 2001), soils (Leininger et al. 2006), and freshwater (Chaban et al. 2006), but their overall species richness in these environments remains unclear. Archaeans are also part of the human gut microflora (Eckburg et al. 2005), and occur in some other animals (Chaban et al. 2006). A survey of methane production across 110 arthropod taxa (Hackstein and Stumm 1994) found evidence for methanogenic archaeans in only a limited set of taxa, including millipedes (Diplopoda), termites (Isoptera), cockroaches (Blattodea), and scarab beetles (Coleoptera: Scarabaeidae). However, they did not find them in other species-rich insect orders, such as Diptera, Hemiptera, Lepidoptera, Orthoptera, or in most beetle families tested (Hymenoptera were not examined). Furthermore, in those arthropod groups in which evidence for the presence of archaeans was found, it was not clear how species-specific they were to their arthropod host species. In summary, we did not include archaeans separately here. However, even if archaean richness were very high (e.g., similar to bacteria), this would alter our estimated Pie of Life, but would only further reinforce our conclusion that global richness was much greater than current estimates and not dominated by insects or animals.

ESTIMATING BACTERIAL RICHNESS

We initially assumed that the many bacterial species associated with insects might have a strong influence on overall bacterial richness. Therefore, to estimate bacterial richness, we first estimated the average number of bacterial gut endosymbionts unique to each arthropod host species. Based on comparisons of closely related species in two of the most species-rich insect orders (Diptera, Hymenoptera), we estimated an average of 10.7 unique bacterial species per arthropod host species (with similar mean values across insect genera; for details see Appendix S5). Furthermore, based on our review of the literature, insects in other species-rich orders (e.g., Coleoptera, Lepidoptera) host similar overall numbers of bacterial species (Appendix S5; Table S3). These rich bacterial biotas are widespread across other animal phyla, including nematodes, chordates, and sponges (Appendix S5).

Given these results, we estimated that each animal species hosts an average of 10.7 unique bacterial species, yielding 1.75 billion bacterial species in total (163.2 million × 10.7). We assumed that the contribution of free-living bacteria to this total was minor in comparison to overall estimated bacterial richness (possibly less than 10 million; Appendix S5). However, it is very important to note that high estimates of free-living bacterial richness would only reinforce our main conclusions about the high richness of bacteria and their dominance of the Pie of Life (see Discussion). We also initially assumed that microsporidian fungi and apicomplexan protists did not host substantial numbers of species-specific bacterial species (being unicellular parasites themselves). Therefore, we did not initially add any projected bacterial richness associated with these two groups.

OVERALL ESTIMATES AND EFFECTS OF CHANGING ASSUMPTIONS

Using the initial estimates for each major group described above, we then summed these estimates across groups to calculate the total number of living species on Earth. We then compared the estimated number of species in each group to this total number to estimate the proprotional richness of each group and the overall Pie of Life.

We also explored the consequences of changing this initial set of assumptions (Table 1), which we hereafter refer to as Sce-

TABLE 1

Projected species number of major taxonomic groups, assuming six cryptic arthropod species per morphology-based species

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Groups	**Parasites with intermediate parasite richness**	**Parasites with reduced parasite richness**	**Parasites with full parasite richness**	**Reduced mite richness, parasites intermediate**
Animals	163.2 million (7%)	163.2 million (11%)	163.2 million (3%)	102 million (7%)
Plants	0.340 million (<0.5%)	0.340 million (<0.5%)	0.340 million (<0.5%)	0.340 million (<0.5%)
Fungi	165.6 million (7%)	165.6 million (11%)	165.6 million (3%)	104.6 million (8%)
Protists	163.2 million (7%)	163.2 million (11%)	163.2 million (3%)	102 million (7%)
Bacteria	1.746 billion (78%)	0.955 billion (66%)	5.264 billion (91%)	1.091 billion (78%)
Total	2.238 billion	1.447 billion	5.756 billion	1.400 billion

Scenario 1 assumes all animal species have a full complement of bacterial, protist, and fungal endosymbionts, even if they are themselves parasites, but that microsporidian fungi and apicomplexan protists have negligible host-specific bacterial species. Scenario 2 assumes that endoparasites have reduced numbers of endosymbiotes themselves (i.e., nematodes have a mean of only one host-specific bacterial species), and that microsporidians and apicomplexans have negligible bacterial species. Scenario 3 assumes that all animal species have a full complement of parasite and endosymbiote species, and that microsporidians and apicomplexans host as many bacterial species as do animal species. Scenario 4 is the same as Scenario 1, but assumes that mites have limited species richness relative to other arthropods (0.25 mites:1 other arthropod species). In this table, all scenarios assume six cryptic arthropod species per morphology-based species (alternative values of two and zero are explored in Tables 2 and 3). Archaean richness is not treated separately.

nario 1. First, we assumed that nematodes living inside insects might have a reduced set of host-specific bacterial species (Scenario 2). Therefore, we assumed each nematode species has (on average) only one unique bacterial species, instead of 10.7. This yielded a total of 873.2 million bacterial species associated with arthropods (81.6 million arthropods × 10.7 bacteria), but only 81.6 million associated with nematodes. This reduced our estimated number of bacterial species to 954.8 million (Table 1).

Second, we assumed that microsporidian fungi and apicomplexan protists hosted as many unique bacterial species as do free-living animal species (Scenario 3). Therefore, we multiplied the number of species of animals (163.2 million), fungi (165.6 million), and protists (163.2 million) by the mean number of host-associated bacterial species estimated for animals (10.7). This yielded a total of 5.264 billion bacterial species (Table 1).

Third, we changed our initial assumption that the number of arthropod-associated mite species is roughly equivalent to the number of other arthropod species (Scenario 4). Instead, we assumed that there are only (on average) 0.25 mites per "other" arthropod species. This specific number (0.25) is arbitrary (and most likely an underestimate; Appendix S5), but is used here for illustrative purposes. We started with our initial estimate of 6.8 million arthropods multiplied by an average of six cryptic species to yield 40.8 million arthropods. We then assumed 0.25 mites per other arthropod species, yielding 10.2 million mite species and a total of 51 million arthropods. This new estimate leads to a lower projected number of nematodes (51 million), animals (102 million), fungi (102 million microsporidians + 2.6 million free-living species, for 104.6 million), protists (102 million), and bacteria (102 million × 10.7 bacteria = 1.091 billion species). Again, we assumed that microsporidians and apicomplexans have limited bacterial richness, as in Scenario 1.

We also explored three additional assumptions, and then addressed their impacts across these four scenarios. First, we assumed that each arthropod species contains (on average) only two cryptic species, instead of six as in our other analyses (see above and Appendix S1 for a discussion of the evidence

TABLE 2

Projected species numbers of major taxonomic groups, assuming two cryptic arthropod species per morphology-based species

Groups	Scenario 1 Parasites with intermediate parasite richness	Scenario 2 Parasites with reduced parasite richness	Scenario 3 Parasites with full parasite richness	Scenario 4 Reduced mite richness, parasites intermediate
Animals	54.4 million (7%)	54.4 million (11%)	54.4 million (3%)	34 million (7%)
Plants	0.340 million (<0.5%)	0.340 million (<0.5%)	0.340 million (<0.5%)	0.340 million (<0.5%)
Fungi	56.8 million (7%)	56.8 million (12%)	56.8 million (3%)	36.4 million (8%)
Protists	54.4 million (7%)	54.4 million (11%)	54.4 million (3%)	34 million (7%)
Bacteria	0.582 billion (78%)	0.318 billion (66%)	1.772 billion (91%)	0.364 billion (78%)
Total	0.748 billion	0.484 billion	1.938 billion	0.468 billion

Estimated numbers of species across major taxonomic groups, under different assumptions (Scenarios 1 to 4), and assuming only two cryptic arthropod species per morphology-based species instead of six (see Table 1 for results assuming six and Table 3 for results assuming zero). See Table 1 for descriptions of Scenarios 1 to 4.

for six versus two cryptic species). This alternative number (two) is somewhat arbitrary, and is merely intended to illustrate the effect of a reduced value (see Table 2). Second, we assumed that there are no cryptic arthropod species (Table 3), although this assumption is demonstrably untrue. Third, we changed our initial assumption that each mite species hosts (on average) one nematode species, like other arthropods. Instead, we assumed that mites harbor negligible nematode richness (see Table 4).

ALTERNATIVE APPROACHES

Our approach is only one among many used to estimate global richness (e.g., Mora et al. 2011; Costello et al. 2012; Locey and Lennon 2016). For example, estimates based on rates of formal taxonomic description of species have estimated much lower global richness (Costello et al. 2012), especially relative to studies of undescribed arthropod richness (Novotny et al. 2002; Basset et al. 2012; Stork et al. 2015). Similar approaches have suggested that parasite richness is much

TABLE 3

Projected species numbers of major taxonomic groups, assuming no cryptic arthropod species

Groups	Scenario 1 Parasites with intermediate parasite richness	Scenario 2 Parasites with reduced parasite richness	Scenario 3 Parasites with full parasite richness	Scenario 4 Reduced mite richness, parasites intermediate
Animals	20.4 million (7%)	20.4 million (10%)	20.4 million (3%)	15.3 million (7%)
Plants	0.340 million (<0.5%)	0.340 million (<0.5%)	0.340 million (<0.5%)	0.340 million (<0.5%)
Fungi	22.8 million (8%)	22.8 million (11%)	22.8 million (3%)	17.7 million (8%)
Protists	20.4 million (7%)	20.4 million (10%)	20.4 million (3%)	15.3 million (7%)
Bacteria	0.218 billion (77%)	0.146 billion (70%)	0.680 billion (91%)	0.164 billion (77%)
Total	0.282 billion	0.209 billion	0.744 billion	0.212 billion

Estimated numbers of species across major taxonomic groups, under different assumptions (Scenarios 1 to 4), and assuming no cryptic arthropod species (see Table 1 for results assuming six cryptic species per morphology-based species, and Table 2 for results assuming two instead). See Table 1 for descriptions of Scenarios 1 to 4.

TABLE 4

Projected species numbers of major taxonomic groups, assuming mites host limited nematode richness, and six cryptic arthropod species per morphology-based species

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Groups	**Parasites with intermediate parasite richness**	**Parasites with reduced parasite richness**	**Parasites with full parasite richness**	**Reduced mite richness, parasites intermediate**
Animals	122.4 million (7%)	122.4 million (9%)	122.4 million (3%)	91.8 million (7%)
Plants	0.340 million (<0.5%)	0.340 million (<0.5%)	0.340 million (<0.5%)	0.340 million (<0.5%)
Fungi	124.8 million (7%)	124.8 million (10%)	124.8 million (3%)	94.2 million (8%)
Protists	122.4 million (7%)	122.4 million (9%)	122.4 million (3%)	91.8 million (7%)
Bacteria	1.310 billion (78%)	0.914 billion (71%)	3.955 billion (91%)	0.982 billion (78%)
Total	1.680 billion	1.284 billion	4.325 billion	1.260 billion

Estimated numbers of species across major taxonomic groups, under different assumptions (Scenarios 1 to 4), and assuming that mites harbor negligible numbers of endoparasitic nematode species (results assuming that all arthropod species host an average of one nematode species are shown in Tables 1–3). See Table 1 for descriptions of Scenarios 1 to 4.

lower than host richness, based on declining parasite description rates over time (Costello 2016). We suggest that the most important parasites for estimating overall global richness are those with potentially species-specific host relationships in the most diverse insect orders (e.g., bacteria, mites, nematodes, apicomplexans, microsporidians). However, the groups that we examine here either did not show such declines (nematodes, microsporidians) or were not examined at all (bacteria, mites, apicomplexans; Costello 2016). Perhaps most importantly, the parasites of insects were simply not addressed (Costello 2016).

On the other hand, we acknowledge that our own estimates may be biased if previous authors primarily studied insects known to host species-specific parasites and endosymbiotes. This might be a possibility in some groups (e.g., mites, nematodes) but seems unlikely for those that have been the subject of systematic, broad-scale surveys (e.g., bacteria). Furthermore, for many eukaryotic parasites (especially internal and unicellular ones, such as nematodes, apicomplexans, and microsporidians), it is unclear how one could predict their presence in particular insect species without detailed studies. That is, it seems unlikely that one could predict that these parasites were present in advance, especially microscopic and internal parasites (i.e., given no external signs of their presence). Therefore, it seems unlikely that the insects studied were chosen because they were known to have these parasites (and that the insects studied are therefore unrepresentative of insects in general). Thus, this source of bias also seems unlikely.

Another approach is to extrapolate local richness (in soils and hosts) to global richness using scaling laws (Locey and Lennon 2016). This approach is promising and gives even higher estimates of global richness than our own (e.g., a trillion species; see Discussion). However, this approach does not estimate the distribution of this richness among taxonomic groups (a major focus of our study) or habitats (e.g., endosymbiotes versus soil). Therefore, it is difficult to address whether and how our results actually differ, beyond the total number of species. Overall, we anticipate that the number of species on Earth (and distribution of these species among groups across the Pie of Life) will become more certain as different approaches converge on similar answers and as the causes of discordance among estimates are illuminated.

RESULTS

THE NUMBER OF SPECIES ON EARTH

We estimated global richness for each major group of organisms, including animals,

plants, fungi, protists, and bacteria (Appendixes S1–S5; Tables S1–S3), under several different scenarios and sets of assumptions (Tables 1–4). Summing the estimates from different groups under these different assumptions, we suggest that Earth most likely has at least a billion species (Table 1), rather than 11 million or fewer (e.g., Mora et al. 2011; Costello et al. 2012; Stork et al. 2015).

We initially estimated ~2.2 billion species (Table 1). However, this depends on several assumptions. We assumed that there are 6.8 million projected species of arthropods based on morphological criteria (Basset et al. 2012), but not including insect-associated mites or morphologically cryptic species revealed by molecular data. Based on a systematic review of studies of molecular species delimitation in arthropods (using coalescent-based methods with multilocus DNA sequence data; Tables S1 and S2; Appendix S1), we estimated that there is an average of six morphologically cryptic species per morphology-based species (but we also explored the effects of assuming only two or zero cryptic species on our overall richness estimates; Tables 2 and 3). This yields a total of 40.8 million non-mite arthropod species (six times 6.8 million). Based on our reviews of the relevant literature (Appendixes S2–S5), we also estimated that each arthropod species hosts (on average) at least one species-specific mite, nematode, apicomplexan protist, and microsporidian fungal species, and a mean of 10.7 bacterial species (Table S3). More precise estimates would be difficult for most groups given the available information. However, these estimates are relatively conservative because they only consider species that are apparently unique to each insect host species.

We initially assumed that endoparasitic animals (e.g., nematodes) contain similar numbers of host-specific bacterial species as do free-living animal species. In contrast, we initially assumed that unicellular parasites (i.e., apicomplexan protists, microsporidian fungi) host trivial bacterial richness. We collectively refer to these latter two assumptions as Scenario 1 (intermediate parasite richness).

We then explored the consequences of changing these assumptions (Table 1). First, we assumed endoparasites have reduced endosymbiote numbers themselves. Specifically, we assumed parasitic nematode species each have only one host-specific bacterial species (on average), and again that microsporidians and apicomplexans have none (Scenario 2; reduced parasite richness). This yields a total of 1.4 billion species summed across all groups. Second, we assumed all animal species have similar bacterial biotas (on average), and that unicellular fungal and protist parasites host roughly as many bacterial species as do animal species (although this seems less likely), yielding a total of ~5.8 billion species (Scenario 3; high parasite richness). Finally, we assumed that mites are much less species-rich than insects, and otherwise followed our initial assumptions about bacterial diversity in endoparasites (i.e., Scenario 1). This scenario (Scenario 4; reduced mite richness) yields 1.4 billion species overall. Thus, across these different scenarios, the projected number of species on Earth remains roughly similar, ranging from ~1 to ~6 billion species (Table 1).

THE NEW PIE OF LIFE

Our estimates of the Pie of Life are also strikingly different from traditional estimates (Figure 1), but are similar across many different estimates of total global richness (Tables 1–4). Pies based on described species numbers have been relatively stable for decades (e.g., Wilson 1992; Roskov et al. 2014). The traditional estimate is ~75% animals (including more than 50% insects) and ~16% plants, ~5% fungi, with only minor contributions from bacteria and protists (e.g., Wilson 1992). Estimates of total richness (including projected undescribed species) that span all groups (e.g., Mora et al. 2011) are even more strongly animal-dominated (about 90% of 11 million projected species), with a similar contribution from fungi (approximately 5%) and a reduced contribution from plants (about 3%). Slices for bacteria and protists are again trivial.

The new Pie of Life estimated here (Figure 1; Table 1) is dominated by bacteria (approximately 70–90% of all species) rather than insects or animals. Furthermore, fungi and protists may be similar in their overall

richness to animals, and the relative richness of all three groups may be quite small (around 3–8%; Table 1). Within animals, insects are only one of three hyperdiverse lineages, along with mites and nematodes.

TESTING ADDITIONAL ASSUMPTIONS

We explored three additional assumptions, and addressed their impacts on total species number across the four scenarios described above. First, we assumed that each morphology-based arthropod species has only two cryptic species (on average), instead of the six assumed in the other analyses. These results are shown in Table 2. This assumption reduces estimates of global richness, decreasing the range of estimates from 1.400–5.756 billion to 0.468–1.938 billion (across the four different scenarios). Nevertheless, global richness remains similar in overall magnitude. The impact on relative richness of major groups (the Pie of Life) is minor (almost all changes of only 1% or less).

Second, we assumed that there are no cryptic arthropod species (despite considerable evidence to the contrary; Appendix S1). The results (Table 3) further lower overall global richness (0.209–0.744 billion), but again remain very large in overall magnitude (hundreds of millions of species) relative to previous estimates. Again, the impact on relative richness of major groups (the Pie of Life) is trivial, with almost all new estimates within 1% of the original estimates.

Third, we changed our initial assumption that each mite species harbors (on average) one nematode species, like other arthropods. Instead, we assumed that mites harbor negligible nematode richness. The impacts of this assumption (Table 4) on our initial estimates of global richness are relatively minor (i.e., changing the range of estimates from 1.400–5.756 billion to 1.260–4.325 billion). Again, the impact on relative richness of major groups is trivial.

DISCUSSION

Our new estimates of global species richness and the Pie of Life are strikingly different from most previous projections. Our results suggest that Earth may have a billion species or more, and that the new Pie of Life is dominated by bacteria (approximately 70–90%) rather than insects or animals. Fungi and protists are similar to animals in their overall richness, and insects are only one of three hyperdiverse animal clades, along with mites and nematodes. Thus, although Haldane quipped that the relative richness of different groups showed that a Creator had an "inordinate fondness for beetles" (Hutchinson 1959:146), our results suggest that a fondness for beetles (and insects in general) was shared with many other groups.

These estimates may seem very surprising. However, they rest on three well-supported observations. First, animal species (especially insects) each have many host-specific bacterial species (Table S3; Appendix S5). Second, insects host many other groups that can be highly specific to a given insect species, including mites, nematodes, microsporidian fungi, and apicomplexan protists. Indeed, some authors commented decades ago that some of these groups had the potential to dramatically increase estimates of global biodiversity (e.g., Ehrlich and Wilson 1991; May 1992). Third, species numbers estimated from morphological data alone underestimate actual richness, even for macroscopic taxa (e.g., Bickford et al. 2007; Adams et al. 2014).

We have also shown that our general conclusions are robust to several assumptions (Tables 1–4). Nevertheless, many uncertainties remain, which should be priorities for future research. These uncertainties are discussed below, and focus on six main areas: the generality of host specificity patterns across arthropods; the impact of free-living species (and additional host-associated clades) in several focal groups; whether host-associated species harbor reduced parastic/endosymbiotic richness themselves; estimating cryptic species diversity in arthropods; the comparability of species across different kingdoms; and the inclusion of viruses.

First, we extrapolated host specificity of many associated groups to all insects or arthropods, even though these groups were sometimes surveyed in relatively few host species (but in species-rich orders). Future

empirical studies should systematically assess species richness and host specificity of these parasite and endosymbiotic groups in closely related insect species in the most diverse insect orders (i.e., Coleoptera, Diptera, Hymenoptera, and Lepidoptera). Our literature-based estimates are conservative (e.g., assuming a single species per host species in some groups), and many groups may be far richer. For example, we focused primarily on parasite, endosymbiote, and other host-associated species that seem to be specific to a given host species. However, species shared among two or more host species may also contribute to overall richness (especially those restricted to only a few host species). Similarly, some groups have additional parasitic or endosymbiotic clades whose host distribution and specificity were not well documented enough for us to confidently include (e.g., in archaeans, protists, and fungi). Furthermore, our estimates of bacterial endosymbiote diversity are from the gut alone, but other organs and tissues may harbor distinct biotas that could multiply bacterial diversity even further (Appendix S5). Finally, many additional factors may influence parasite/endosymbiote richness that were not directly addressed here, such as host body size or geographic range size (e.g., Poulin and Morand 2004; Kamiya et al. 2014). However, most of our estimates are minimum estimates (e.g., for apicomplexans, microsporidians, mites, and nematodes), which should not change with these factors. Similarly, estimates of bacterial richness are broadly similar across a wide range of insect hosts (Appendix S5). Overall, we suggest that incorporating these other factors might increase estimates of richness of these parasitic and endosymbiotic groups, but should not overturn our minimum estimates.

Second, in some groups we did not include estimates of their free-living diversity, especially in soils (e.g., mites, nematodes, protists, and bacteria). We made this choice because it is often difficult to extrapolate this diversity to the global scale. We readily acknowledge that the free-living biodiversity of soils may also be extensive, and its inclusion might change our estimates somewhat. However, additions of a few million species, as for soil fungi (Tedersoo et al. 2014), would have little impact on our main results (Table 1). Some previous studies have also estimated much higher diversity of soil bacteria (e.g., Dykhuizen 1998), although these estimates have been controversial (e.g., Schloss and Handelsman 2004). More recently, Locey and Lennon (2016) estimated that there are ~1 trillion microbial species, but did not address where these species occur (e.g., free-living versus endosymbiotic) or to what groups they belonged (e.g., archaeans, bacteria, fungi, protists). Most importantly for our study, estimating much larger numbers of free-living species (or more parasitic or endosymbiotic species) from groups such as mites, nematodes, protists, fungi, archaeans, and bacteria would only reinforce our two most fundamental conclusions: first, that overall species richness is much higher than most previous estimates and, second, that the Pie of Life is not dominated by insects or animals.

A third key area of uncertainty is whether host-associated species themselves have reduced parasite or endosymbiote diversity. For example, do nematodes inside insects have the same mean bacterial richness as free-living nematodes? Do unicellular parasites (e.g., apicomplexans, microsporidians) have reduced bacterial richness, given their size or occurrence inside other organisms? We explored how different answers to these questions might influence our results, and found that our major conclusions were robust (Tables 1–4). Nevertheless, these questions should be thoroughly explored with focused empirical studies. Most importantly, we explicitly address these uncertainties here, but they are clearly relevant to all studies that estimate the overall species richness of life on Earth.

Fourth, we illustrate how estimates of overall species richness on Earth depend on the relative numbers of morphologically distinct and cryptic arthropod species. We initially assumed six cryptic species per estimated morphology-based species, but we have shown that our major conclusions remain supported regardless of whether we assume that there are two or six cryptic arthropod species per projected morphology-based spe-

cies, or even zero (although this latter number seems very unlikely; see Appendix S1). We show that assuming fewer cryptic arthropod species reduces overall richness (but with the lowest estimates still in the hundreds of millions) and has negligible impact on our estimates of the Pie of Life (Tables 2 and 3). We also assumed that these cryptic species each have distinct parasite and endosymbiote species. We think that the latter is a very reasonable assumption (and therefore we did not explore it explicitly). After all, if cryptic host species are geographically isolated from each other (see below), then their parasites and endosymbiotes should be also. However, more explicit testing would be desirable. Conversely, parasites and endosymbiotes might speciate through geographic isolation even when their hosts do not (Wong et al. 2013). This process might increase the richness of these groups even further.

In a similar vein, the baseline projections of global insect richness used here are based (in part) on the association of insects with plants (Novotny et al. 2002; Basset et al. 2012; Stork et al. 2015), but these estimates do not necessarily incorporate morphologically cryptic species associated with different plant hosts. Our estimates of morphologically cryptic arthropod species are primarily associated with single morphology-based species that occur in different geographic regions (where different regions have distinct cryptic species), rather than on different host plant species (Appendix S1). However, there might also be morphologically cryptic insect species on different host plants. For example, in one well-known study, a single morphology-based butterfly species was estimated to contain 10 morphologically cryptic species that occur in sympatry, each utilizing largely distinct host plant species (Hebert et al. 2004). However, this example may be more extreme than typical. Similarly, an insect species that uses different host plants locally might have different morphologically cryptic parasitoid species on different host plants (e.g., wasps; Forbes et al. 2009), even if the host insect has not speciated. If such cases were frequent, it could increase insect richness further (at least for herbivorous insects and their parasitoids), presumably with cascading increases in richness in their parasites and endosymbiotes. This possibility adds support to our initial assumption that there could be many cryptic arthropod species for each currently described species, and thus many cryptic species for each species projected to exist based on morphological evidence (e.g., Novotny et al. 2002; Basset et al. 2012; Stork et al. 2015). However, we emphasize that the cryptic species considered in this study were inferred from molecular evidence showing that populations in different parts of the geographic range of morphology-based species are actually distinct species. Therefore, the global number of cryptic species cannot be estimated solely from studies of cryptic species at a single geographic location (e.g., Hebert et al. 2004).

Fifth, a broader question is whether species richness can be compared between different groups at all. However, there is no reason why this should be an issue for our study alone, and not other studies on this topic, especially those including both eukaryotes and prokaryotes (e.g., Mora et al. 2011; Locey and Lennon 2016). Perhaps the most important issue is that an asexual bacterial species may not be identical to a sexual plant and animal species, in which species limits and speciation are (in the end) typically based on inferences about gene flow and reproductive isolation (e.g., Coyne and Orr 2004). Nevertheless, there is genetic exchange among closely related bacterial individuals through homologous recombination (e.g., Fraser et al. 2007; Shapiro et al. 2012). Therefore, recombination may act as a cohesive force among conspecific bacterial individuals, just as in sexual species (e.g., Fraser et al. 2007; Shapiro et al. 2012). Many bacterial species also appear to be comparable to many sexual species based on additional critera, such as: homogenization of conspecific individuals through ecological similarity and natural selection (in addition to recombination); phenotypic clustering of conspecific individuals; and adaptive divergence among species (Cohan 2002; Vos 2011; Rosselló-Móra and Amann 2015). Species concepts and delimitation in bacteria are subjects

of ongoing debate and discussion (e.g., Cohan 2002; Vos 2011; Rosselló-Móra and Amann 2015), but this is true for all other organisms as well. After all, there is considerable variation in sexual systems (and species isolation mechanisms) among different plant species, among animal species, and between animals and plants (e.g., Coyne and Orr 2004). In summary, we argue that bacterial species should be broadly comparable to those in other groups.

Our estimates of bacterial species richness (i.e., within insect hosts) are based on a standard and highly conservative criterion for bacterial species delimitation (i.e., clustering of 16S sequences at the approximately 97% identity level; see Appendix S5). Other criteria might actually show bacterial richness to be considerably higher (e.g., Hong et al. 2009). Using this minimal criterion of ~3% sequence divergence, some estimates indicate that most bacterial species may be millions of years old (Ochman et al. 1999), and potentially more divergent than closely related plant and animal species. Thus, the much higher richness of bacterial species estimated here (relative to many previous estimates) should not be an artifact related to these bacterial species being insufficiently distinct from each other (or somehow less "real" than plant, animal, or fungal species).

Finally, our study excludes viruses because their status as living organisms is unclear (Moreira and López-García 2009). Nevertheless, some authors suggest that each bacterial species might host 10 or more unique virus species (Rowher 2003). Therefore, if they were included, global biodiversity might extend into the tens of billions, and the Pie of Life would then be dominated by viruses (more than 90%).

Conclusions

In summary, we provide new estimates of global species richness and the Pie of Life that are dramatically different from traditional estimates. Clearly, our specifc results are subject to considerable uncertainty, and we present broad ranges of global richness estimates rather than a single value. Nevertheless, we argue that all future estimates of global biodiversity should incorporate morphologically cryptic species and the diversity of clades and species hosted by arthropods. If they do, then future estimates of global richness and the Pie of Life should remain orders of magnitude higher than most current estimates, and should be dominated by non-eukaryotic groups. Our results also imply that parasitic, endosymbiotic, and similar relationships between species might be fundamental drivers of the overall richness of life on Earth.

acknowledgments

We thank D. Walter, H. Proctor, O. Seeman, and P. Stock for useful advice on mites and nematodes. Brendan B. Larsen and Elizabeth C. Miller were supported by Graduate Research Fellowships from the U.S. National Science Foundation.

REFERENCES

Adams M., Raadik T. A., Burridge C. P., Georges A. 2014. Global biodiversity assessment and hyper-cryptic species complexes: more than one species of elephant in the room? *Systematic Biology* 63:518–533.

Basset Y., Cizek L., Cuénoud P., et al. 2012. Arthropod diversity in a tropical forest. *Science* 338:1481–1484.

Bickford D., Lohman D. J., Sodhi N. S., Ng P. K. L., Meier R., Winker K., Ingram K. K., Das I. 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution* 22:148–155.

Camargo A., Morando M., Avila J. L., Sites J. W. Jr. 2012. Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwinii* complex (Squamata: Liolaemidae). *Evolution* 66:2834–2849.

Carstens B. C., Pelletier T. A., Reid N. M., Satler J. D. 2013. How to fail at species delimitation. *Molecular Ecology* 22:4369–4383.

Chaban B., Ng S. Y. M., Jarrell K. F. 2006. Archaeal habitats—from the extreme to the ordinary. *Canadian Journal of Microbiology* 52:73–116.

Cohan F. M. 2002. What are bacterial species? *Annual Review of Microbiology* 56:457–487.

Costello M. J. 2016. Parasite rates of discovery, global species richness and host specificity. *Integrative and Comparative Biology* 56:588–599.

Costello M. J., Wilson S., Houlding B. 2012. Predicting total global species richness using rates of species description and estimates of taxonomic effort. *Systematic Biology* 61:871–883.

Coyne J. A., Orr H. A. 2004. *Speciation*. Sunderland (Massachusetts): Sinauer Associates.

Curtis T. P., Sloan W. T., Scannell J. W. 2002. Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences of the United States of America* 99:10494–10499.

Dykhuizen D. E. 1998. Santa Rosalia revisited: why are there so many species of bacteria? *Antonie van Leeuwenhoek* 73:25–33.

Eckburg P. B., Bik E. M., Bernstein C. N., Purdom E., Dethlefsen L., Sargent M., Gill S. R., Nelson K. E., Relman D. A. 2005. Diversity of the human intestinal microbial flora. *Science* 308:1635–1638.

Ehrlich P. R., Wilson E. O. 1991. Biodiversity studies: science and policy. *Science* 253:758–762.

Erwin T. L. 1982. Tropical forests: their richness in Coleoptera and other arthropod species. *Coleopterist's Bulletin* 36:74–75.

Forbes A. A., Powell T. H. Q., Stelinski L. L., Smith J. J., Feder J. L. 2009. Sequential sympatric speciation across trophic levels. *Science* 323:776–779.

Fraser C., Hanage W. P., Spratt B. G. 2007. Recombination and the nature of bacterial speciation. *Science* 315:476–480.

Hackstein J. H. P., Stumm C. K. 1994. Methane production in terrestrial arthropods. *Proceedings of the National Academy of Sciences of the United States of America* 91:5441–5445.

Hebert P. D. N., Penton E. H., Burns J. M., Janzen D. H., Hallwachs W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America* 101:14812–14817.

Hong S., Bunge J., Leslin C., Jeon S., Epstein S. S. 2009. Polymerase chain reaction primers miss half of rRNA microbial diversity. *ISME Journal* 3:1365–1373.

Hutchinson G. E. 1959. Homage to Santa Rosalia or why are there so many kinds of animals? *American Naturalist* 93:145–159.

Joppa L. N., Roberts D. L., Pimm S. L. 2011. How many species of flowering plants are there? *Proceedings of the Royal Society B: Biological Sciences* 278:554–559.

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

Karner M. B., DeLong E. F., Karl D. M. 2001. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* 409:507–510.

Lambshead P. J. D. 1993. Recent developments in marine benthic biodiversity research. *Oceanis* 19:5–24.

Leininger S., Urich T., Schloter M., Schwark L., Qi J., Nicol G. W., Prosser J. I., Schuster S. C., Schleper C. 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 442:806–809.

Locey K. J., Lennon J. T. 2016. Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences of the United States of America* 113:5970–5975.

May R. M. 1992. How many species inhabit the Earth? *Scientific American* 267:18–24.

Mora C., Tittensor D. P., Adl S., Simpson A. G. B., Worm B. 2011. How many species are there on Earth and in the ocean? *PLOS Biology* 9:e1001127.

Moreira D., López-García P. 2009. Ten reasons to exclude viruses from the tree of life. *Nature Reviews Microbiology* 7:306–311.

Novotny V., Basset Y., Miller S. E., Weiblen G. D., Bremer B., Cizek L., Drozd P. 2002. Low host specificity of herbivorous insects in a tropical forest. *Nature* 416:841–844.

Ochman H., Elwyn S., Moran N. A. 1999. Calibrating bacterial evolution. *Proceedings of the National Academy of Sciences of the United States of America* 96:12638–12643.

Poulin R., Morand S. 2004. *Parasite Biodiversity*. Washington (DC): Smithsonian Books.

Rabosky D. L., Matute D. R. 2013. Macroevolutionary speciation rates are decoupled from the evolution of intrinsic reproductive isolation in *Drosophila* and birds. *Proceedings of the National Academy of Sciences of the United States of America* 110:15354–15359.

Roskov Y. et al. 2014. Species 2000 and ITIS Catalogue of Life, 29 October 2014. Leiden (The Netherlands): Species 2000, 2014. Available at http://www.catalogueoflife.org/col.

Rosselló-Móra R., Amann R. 2015. Past and future species definitions for *Bacteria* and *Archaea*. *Systematic and Applied Microbiology* 38:209–216.

Rowher F. 2003. Global phage diversity. *Cell* 113:141.

Scheffers B. R., Joppa L. N., Pimm S. L., Laurance W. F. 2012. What we know and don't know about Earth's missing biodiversity. *Trends in Ecology and Evolution* 27:501–510.

Schloss P. D., Handelsman J. 2004. Status of the microbial census. *Microbiology and Molecular Biology Reviews* 68:686–691.

Shapiro B. J., Friedman J., Cordero O. X., Preheim S. P., Timberlake S. C., Szabó G., Polz M. F., Alm E. J. 2012. Population genomics of early events in the ecological differentiation of bacteria. *Science* 336:48–51.

Stork N. E., McBroom J., Gely C., Hamilton A. J. 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.

Sukumaran J., Knowles L. L. 2017. Multispecies coalescent delimits structure, not species. *Proceedings of the*

National Academy of Sciences of the United States of America 114:1607–1612.

Tedersoo L., Bahram M., Põlme S., et al. 2014. Global diversity and geography of soil fungi. *Science* 346:1256688.

Vos M. 2011. A species concept for bacteria based on adaptive divergence. *Trends in Microbiology* 19:1–7.

Wiens J. J. 2004. What is speciation and how should we study it? *American Naturalist* 163:914–923.

Wiens J. J., Engstrom T. N., Chippindale P. T. 2006. Rapid diversification, incomplete isolation, and the "speciation clock" in North American salamanders (genus *Plethodon*): testing the hybrid swarm hypothesis in rapid radiation. *Evolution* 60:2585–2603.

Wilson E. O. 1992. *The Diversity of Life*. Cambridge (Massachusetts): Belknap Press of Harvard University Press.

Wong A. C.-N., Chaston J. M., Douglas A. E. 2013. The inconstant gut microbiota of *Drosophila* species revealed by 16S rRNA gene analysis. *ISME Journal* 7: 1922–1932.

Yang Z., Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America* 107:9264–9269.

Handling Editor: Daniel E. Dykhuizen

APPENDIX S1

Estimating the Number of Cryptic Arthropod Species

Cryptic species (as defined here) are present when a named, morphologically defined species is found to contain two or more species that are distinct (e.g., based on molecular data), but which are not necessarily morphologically distinguishable. Cryptic species have been called the "elephant in the room" in determining global species richness (Adams et al. 2014). Yet, studies estimating global richness vary in the extent to which they explicitly consider cryptic species, from including estimates of cryptic species richness based on expert opinion (Appeltans et al. 2012) to a brief discussion of the topic (Scheffers et al. 2012; Stork et al. 2015) to no mention at all (Mora et al. 2011; Costello et al. 2013).

Here, we attempt to quantitatively estimate the mean number of cryptic species per morphology-based arthropod species, in order to refine our estimates of global species richness. We focus only on arthropods because for other animal groups, even if there were numerous cryptic species per described species, their richness would still be inconsequential relative to that for arthropods. For example, if every named vertebrate species contained an average of 10 undescribed cryptic species, this would still add up to less than a million species.

To estimate the number of cryptic species, we only considered cryptic species as distinct if they were strongly supported using a rigorous method of species delimitation and multilocus genetic data (i.e., we did not simply count species estimated through DNA barcoding of a single mitochondrial gene). We conducted a literature search for studies using the multilocus coalescent model implemented in Bayesian Phylogenetics and Phylogeography (BPP; Yang and Rannala 2010). We chose this method because it is quickly becoming a standard approach for species delimitation (Carstens et al. 2013), and because it has outperformed other species delimitation methods in accurately identifying evolutionary lineages in simulated and empirical data (Camargo et al. 2012). Unlike "discovery" delimitation methods, which assign individuals to species groups without a priori information, BPP is a "validation" delimitation method, which requires a resolved phylogeny of putative species as input along with multilocus sequence data. The method then tests the support for the species status for each of the nodes in the given phylogeny. This validation aspect of BPP is particularly important to our question here, because it can provide an objective test of the hypothesis that a described species contains one or more cryptic lineages (meaning a hypothesized cryptic species is either validated with high support or it is not). Therefore, it is possible for putative cryptic species to be statistically unsupported, which should provide more conservative estimates than the number of putative species inferred solely from an author's interpretation of the data.

We used Google Scholar to search for studies of arthropods that used BPP. We used the following Google Scholar query: "cryptic," with the exact phrase "Bayesian species," at least one of the words "delimitation" or "delineation," and the years 2013–2015 (we assumed applications of BPP would be limited immediately after its publication). The search was conducted on December 20, 2015. This query yielded 284 results, from which we only selected studies using BPP on arthropods (20 total). We did not include a study on parasitic mites (Bochkov et al. 2014) because estimates of host-specific mites are included separately in our analyses of arthropod richness (Appendix S2).

In each study, we counted the number of cryptic species delimited with high support for

each described species included in the BPP analysis (ignoring results of other species delimitation analyses in the study). We did not include new species discovered using the BPP method that were not originally identified as part of a named species because our goal was to calculate the ratio of cryptic to described species. For this reason, we excluded three studies because the information provided did not allow us to distinguish between cryptic species within a described species from newly discovered species, given the unclear taxonomic affinity of the putative species (Parmakelis et al. 2013; Derkarabetian and Hedin 2014; Esmaeili-Rineh et al. 2015). Each study considered used a combination of mitochondrial and nuclear loci. Furthermore, the geographic ranges of the described species that were included were broadly sampled (i.e., multiple populations across the species range). Importantly, almost all cryptic species that were delimited were based on subdividing the geographic range of a described species. The sole exception to our focus on cryptic geographically distinct species was a study on host-related cryptic species of parasitoid wasps (Hamback et al. 2013).

We ultimately analyzed 16 studies, and found the mean number of cryptic molecular species supported by BPP per described species to be 5.9 (Table S1). This is the mean of the 16 study-specific means (Table S2). There was little difference in numbers between insects and non-insect arthropods (Table S1). We apply the number six (5.9 rounded up) to estimate the overall number of arthropod species in the main text (Table 1).

We also considered alternative values of two and zero cryptic species in calculating global richness (Tables 2 and 3). Although zero cryptic species is unlikely, our estimated value of six might also be an overestimate, for several reasons.

First, our results may be biased if authors only published studies that revealed cryptic species. However, the impact of publication bias is unclear, because many studies using BPP were primarily concerned with broader questions of phylogeography, and not simply the discovery of cryptic species. Therefore, they do report negative results: cases in which potential cryptic species were not supported (e.g., Hedin 2015; Toussaint et al. 2015).

Second, recent simulations suggest that the BPP method might overestimate the number of species when there is a time lag between genetic fragmentation and completion of the speciation process (Sukumaran and Knowles 2017). On the other hand, the criterion used in the simulations may underestimate species richness.

Third, it is possible that our estimated value of 5.9 cryptic species is biased to be high because some studies that used BPP were focused on "species complexes" (i.e., Zhang et al. 2014). Therefore, these might be groups in which undescribed species (or other taxonomic issues) were suspected, and so might have more cryptic species than a completely random sample of arthropod species. On the other hand, numerous species might be too poorly known (morphologically or geographically) to be recognized as "species complexes" in the first place, especially since the vast majority of estimated species richness in insects is undescribed. In addition, studies that densely sampled each species across its geographic range tended to find more cryptic species than studies with less dense sampling per species (i.e., Zhang et al. 2014 versus Toussaint et al. 2015).

Fourth, there are alternative ways to estimate the number of cryptic species from our data. Specifically, one could simply take the total number of lineages supported by BPP and divide this by the total number of described species included in all of these studies, rather than the mean of 16 study-specific means. This would yield 2.00 cryptic species per described species (82 versus 41) for insects, 3.50 for non-insect arthropods (84 versus 24), and 2.55 for arthropods overall (166 versus 65). However, this approach is somewhat problematic. Specifically, the results for insects using this approach are dominated by a single study that included 32 described species but found relatively few cryptic species among them, presumably because of decreased sampling effort per described species (Toussaint et al. 2015). Excluding this study yields a mean of five cryptic species per described insect species, similar to the mean value across studies. Therefore, the average across studies ameliorates such potential artifacts.

These issues are difficult to resolve at the present time, and have been the subject of considerable debate. For example, Pfenninger and Schwenk (2007) argued that cryptic species are distributed homogenously among higher taxa when correcting for study effort, whereas Pérez-Ponce de León and Poulin (2016) found conflicting results a decade later. In this study, we consider six versus two or zero cryptic species per named species (on average) to estimate the impact of this uncertainty on global species richness (Tables 1, 2, and 3 in main text, respectively).

Additionally, several authors have suggested that synonyms (multiple names corresponding to the same biological species) might counteract

the additions to global species richness from cryptic species (Mora et al. 2011; Appeltans et al. 2012; Pimm 2012). Authors differ in their view on the role of synonyms in estimating global species richness, from minor (Appeltans et al. 2012) to major (Scheffers et al. 2012; Stork et al. 2015). However, none of these studies have directly claimed that there are more synonyms than cryptic species. Our survey takes potential synonymy into account because BPP analyses could also reveal that the number of named species was overestimated (i.e., if molecular data did not support morphologically defined species as distinct). For example, in the study by Hedin (2015), several species named based on morphological evidence are not supported as distinct by BPP. In the end, our result of 5.9 cryptic species per morphologically based arthropod species suggests that additions to global species richness from cryptic species may vastly outweigh subtractions due to synonymy (at least for arthropods).

REFERENCES

Adams M., Raadik T. A., Burridge C. P., Georges A. 2014. Global biodiversity assessment and hyper-cryptic species complexes: more than one species of elephant in the room? *Systematic Biology* 63:518–533.

Appeltans W., Ahyong S. T., Anderson G., et al. 2012. The magnitude of global marine species diversity. *Current Biology* 22:2189–2202.

Bochkov A. V., Klimov P. B., Hestvik G., Saveljev A. P. 2014. Integrated Bayesian species delimitation and morphological diagnostics of chorioptic mange mites (Acariformes: Psoroptidae: *Chorioptes*). *Parasitology Research* 113:2603–2627.

Camargo A., Morando M., Avila J. L., Sites J. W. Jr. 2012. Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwinii* complex (Squamata: Liolaemidae). *Evolution* 66:2834–2849.

Carstens B. C., Pelletier T. A., Reid N. M., Satler J. D. 2013. How to fail at species delimitation. *Molecular Ecology* 22:4369–4383.

Costello M. J., May R. M., Stork N. E. 2013. Can we name Earth's species before they go extinct? *Science* 339:413–416.

Derkarabetian S., Hedin M. 2014. Integrative taxonomy and species delimitation in harvestmen: a revision of the western North American genus *Sclerobunus* (Opiliones: Laniatores: Travunioidea). *PLOS ONE* 9:e104982.

Dickey A. M., Kumar V., Hoddle M. S., Funderburk J. E., Morgan J. K., Jara-Cavieres A., Shatters R. G. Jr., Osborne L. S., McKenzie C. L. 2015. The *Scirtothrips dorsalis* species complex: endemism and invasion in a global pest. *PLOS ONE* 10:e0123747.

Esmaeili-Rineh S., Sari A., Delić T., Moškrič A., Fišer C. 2015. Molecular phylogeny of the subterranean genus *Niphargus* (Crustacea: Amphipoda) in the Middle East: a comparison with European niphargids. *Zoological Journal of the Linnean Society* 175:812–826.

Hambäck P. A., Weingartner E., Ericson L., Fors L., Cassel-Lundhagen A., Stenberg J. A., Bergsten J. 2013. Bayesian species delimitation reveals generalist and specialist parasitic wasps on *Galerucella* beetles (Chrysomelidae): sorting by herbivore or plant host. *BMC Evolutionary Biology* 13:92.

Hedin M. 2015. High-stakes species delimitation in eyeless cave spiders (*Cicurina*, Dictynidae, Araneae) from central Texas. *Molecular Ecology* 24:346–361.

Hedin M., Carlson D., Coyle F. 2015. Sky island diversification meets the multispecies coalescent–divergence in the spruce-fir moss spider (*Microhexura montivaga*, Araneae, Mygalomorphae) on the highest peaks of southern Appalachia. *Molecular Ecology* 24:3467–3484.

Hsieh C.-H., Ko C.-C., Chung C.-H., Wang H.-Y. 2014. Multilocus approach to clarify species status and the divergence history of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. *Molecular Phylogenetics and Evolution* 76:172–180.

Mora C., Tittensor D. P., Adl S., Simpson A. G. B., Worm B. 2011. How many species are there on Earth and in the ocean? *PLOS Biology* 9:e1001127.

Murphy N. P., King R. A., Delean S. 2015. Species, ESUs or populations? Delimiting and describing morphologically cryptic diversity in Australian desert spring amphipods. *Invertebrate Systematics* 29:457–467.

Opatova V., Arnedo M. A. 2014. Spiders on a hot volcanic roof: colonisation pathways and phylogeography of the Canary Islands endemic trap-door spider *Titanidiops canariensis* (Araneae, Idiopidae). *PLOS ONE* 9: e115078.

Parmakelis A., Kotsakiozi P., Stathi I., Poulikarakou S., Fet V. 2013. Hidden diversity of *Euscorpius* (Scorpiones: Euscorpiidae) in Greece revealed by multilocus species-delimitation approaches. *Biological Journal of the Linnean Society* 110:728–748.

Pedraza-Lara C., Barrientos-Lozano L., Rocha-Sánchez A. Y., Zaldívar-Riverón A. 2015. Montane and coastal species diversification in the economically important Mexican grasshopper genus *Sphenarium* (Orthoptera: Pyrgomorphidae). *Molecular Phylogenetics and Evolution* 84:220–231.

Pérez-Ponce de León G., Poulin R. 2016. Taxonomic distribution of cryptic diversity among metazoans: not so homogeneous after all. *Biology Letters* 12:20160371.

Pfenninger M., Schwenk K. 2007. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology* 7:121.

Pimm S. L. 2012. Biodiversity: not just lots of fish in the sea. *Current Biology* 22:R996–R997.

Satler J. D., Carstens B. C., Hedin M. 2013. Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, *Aliatypus*). *Systematic Biology* 62:805–823.

Scheffers B. R., Joppa L. N., Pimm S. L., Laurance W. F. 2012. What we know and don't know about Earth's missing biodiversity. *Trends in Ecology and Evolution* 27:501–510.

Stork N. E., McBroom J., Gely C., Hamilton A. J. 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.

Su Y.-C., Wang J.-F., Villanueva R. J. T., Nuñeza O. M., Lin C.-P. 2014. Hopping out of Mindanao: Miocene-Pliocene geological processes and cross-island dispersal as major drivers of diversity for Philippine treehoppers. *Journal of Biogeography* 41:1277–1290.

Sukumaran J., Knowles L. L. 2017. Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences of the United States of America* 114:1607–1612.

Toussaint E. F. A., Morinière J., Müller C. J., Kunte K., Turlin B., Hausmann A., Balke M. 2015. Comparative molecular species delimitation in the charismatic Nawab butterflies (Nymphalidae, Charaxinae, *Polyura*). *Molecular Phylogenetics and Evolution* 91:194–209.

Wachter G. A., Muster C., Arthofer W., Raspotnig G., Föttinger P., Komposch C., Steiner F. M., Schlick-Steiner B. C. 2015. Taking the discovery approach in integrative taxonomy: decrypting a complex of narrow-endemic Alpine harvestmen (Opiliones: Phalangiidae: *Megabunus*). *Molecular Ecology* 24:863–889.

Wade E. J., Hertach T., Gogala M., Trilar T., Simon C. 2015. Molecular species delimitation methods recover most song-delimited cicada species in the European *Cicadetta montana* complex. *Journal of Evolutionary Biology* 28:2318–2336.

Yang Z., Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America* 107:9264–9269.

Zhang F., Yu D., Luo Y., Ho S. Y. W., Wang B., Zhu C. 2014. Cryptic diversity, diversification and vicariance in two species complexes of *Tomocerus* (Collembola, Tomoceridae) from China. *Zoologica Scripta* 43:393–404.

Zhang Y., Li S. 2013. Ancient lineage, young troglobites: recent colonization of caves by *Nesticella* spiders. *BMC Evolutionary Biology* 13:183.

Zhang Y., Li S. 2014. A spider species complex revealed high cryptic diversity in South China caves. *Molecular Phylogenetics and Evolution* 79:353–358.

APPENDIX S2

Estimating Animal Richness

Overview

Current estimates of species numbers suggest that arthropods, especially insects, are the most species-rich clade of animals (and living things). The current catalog of living, described species suggests that there are ~1 million described species of insects, out of 1.6 million described species in total across the Tree of Life (Roskov et al. 2014). Thus, 62% of all living species are insects based on these values. We suggest that the proportional richness of insects may be overestimated. Specifically, groups that parasitize insects (or are otherwise intimately associated with them) may occupy a larger slice of the Pie of Life. We address these groups in turn.

Phylum Arthropoda

insects

The number of currently described insect species is ~1 million (Roskov et al. 2014). There is no question that there are many more undescribed species, but the specific number of undescribed species has varied among different estimates. However, in recent years, several estimates have begun to converge toward similar numbers. For example, Hamilton et al. (2011) and Basset et al. (2012) tentatively agreed on a number around 6.1 million for global tropical arthropod species richness. Importantly, "arthropod" is not just insects, and also includes mites, but the contribution of mites is low in these estimates. Note that Hamilton et al. (2010) included an error that caused diversity to be underestimated by nearly twofold; this was corrected by Hamilton et al. (2011). Hamilton et al. (2013) noted that there was considerable uncertainty in this estimate.

Stork et al. (2015) recently showed that four different methods for estimating beetle richness converged on a similar mean estimate of 1.5 million beetle species (both described and undescribed), with a relatively narrow range (0.9–2.1 million). They used relationships between

plant and beetle diversity to estimate total global insect richness, finding that this was from 5.4–7.2 (midpoint = 6.3). They assumed that there is one beetle species per 3.9 species of other terrestrial arthropods, to estimate a total of 5.9–7.4 terrestrial arthropod species (mean among estimates of 6.8 million). However, we do not use this estimate directly, given that that this most likely underestimates the diversity of mites that utilize other arthropods as hosts (see next section).

MITES

Almost all organisms that are larger than mites are colonized by mites (Walter and Proctor 2013), including plants, vertebrates, molluscs, annelids, and arthropods (including larger mites). A single host species may harbor a very large number of mite species, potentially from many different mite families. Levels of host specificity are not always clear, but Walter and Proctor (2013) noted that there has been a tendency toward host specificity in the diversification of symbiotic mites in association with many species-rich host clades, including hymenopterans (ants, bees, wasps), coleopterans (beetles), cockroaches, orthopterans, hemipterans, myriapods (centipedes and millipedes), squamates (snakes and lizards), birds, and mammals, conifers, and angiosperms.

MITES ON INSECTS

Mites often occur on insects, and they may do so in a remarkable variety of locations and taxa (Walter and Proctor 2013), such as the tracheae of honey bees (Hymenoptera), the stink glands of stink bugs (Hemiptera), and the ears of noctuid moths (Lepidoptera). Not all associations between mites and hosts are clearly parasitic, including many mites that use other arthropods for transportation (phoresy; Walter and Proctor 2013). Many mite species have potentially commensal relationships with insects, or relationships where the costs and benefits are not obvious. One genus from one family occurs on the antennae of army ants, whereas another genus from another family occurs on their feet (Walter and Proctor 2013).

Walter and Proctor (2013) reviewed associations between mites and their hosts across different groups of mites. They noted that mesostigmatans (in Parasitiformes) are particularly diverse on insects and other arthropods, including the most species-rich insect orders (Coleoptera, Diptera, Hymenoptera, and Lepidoptera). These mites utilize many families in each of these orders, including at least 24 coleopteran families, 19 dipteran families, eight hymenopteran families, and 41 lepidopteran families (review in Hunter and Rosario 1988). Mestostigmatans also utilize non-insect arthropods, including centipedes, millipedes, scorpions, spiders, and crustaceans (Hunter and Rosario 1988). Furthermore, mesostigmatans are not the only mites occurring on insects (Walter and Proctor 2013). For example, Prostigmata (in Acariformes) are also frequent parasites on arthropods, including all major orders of insects and arachnids. Parasitengonina includes water mites that parasitize most major groups of freshwater insects (including dipterans, odonates, hemipterans, and coleopterans, but not ephemeropterans or megalopterans; Walter and Proctor 2013).

There can also be large numbers of mite species and families on each insect host species. Honey bees alone are host to 34 mite species (De Jong et al. 1982). Up to 55 families of mites are associated with a single army ant species, and even a temperate ant species may host more than 30 species of mites (Walter and Proctor 2013). Examples of high mite diversity on beetles include 13 mite species from three families from the scarabaeid species *Heliocopris japetus*, 19 species from four families from the scarabaeid *Copris hispanus*, and 13 species from eight families from the passalid *Odontotaenius disjunctus* (review in Hunter and Rosario 1988). Even very small insects can have many mite species that are intimately associated with them (i.e., approximately 5 mm carabid beetles of the genus *Sericoda*; Beaulieu et al. 2008).

Mites on insects can exhibit a diverse range of host specificities: a single insect host species can harbor dozens of mite species, but a single mite species can also utilize multiple hosts (Walter and Proctor 2013). Unfortunately, the data that would provide a strong basis for estimating a ratio of insect to mite diversity across all insects seem to be lacking (i.e., published surveys of the entire mite faunas of closely related insect species across many orders), and it is possible that published surveys have focused only on those insect taxa known to have many mites. Nevertheless, given that insect-associated mites can be highly diverse on individual host species (i.e., potentially compensating for those host species that might lack host-specific mite species altogether), that they are widespread across the most diverse insects orders (e.g., Coleoptera, Diptera, Hymenoptera, Lepidoptera), and that many mite species can be quite host specific, we tentatively assume that there is (on average)

at least one mite species per insect species (for specific studies supporting this ratio see Schwartz et al. 1998; Salmane and Telnov 2009; Knee et al. 2012, 2013; specifically these four studies respectively estimated ratios of beetle species to mite species of 4:7, 38:36, 51:36, and 30:33 for a mean ratio across studies of 1.13 or a combined ratio, based on species and not separate studies, of 123:112 or 0.91).

This ratio may be an overestimate, and we dealt with this possibility in two ways. First, we did not include mites inhabiting other animals or plants as part of our tally (even though the number of mite species on plants could be very large, see below). This was intended to counterbalance our potential overestimation for insect-associated mites. Second, we performed a set of analyses in which we assumed that the ratio of insect to mite diversity was much smaller, only 1:0.25 instead of 1:1.

MITES ON OTHER ANIMALS

Mites are known to be particularly diverse on the skins, fur, and feathers of birds and mammals (Walter and Proctor 2013). We make the simplifying assumption that mites on vertebrates are relatively well studied (given that thousands of species have already been described). Moreover, even if each bird and mammal species had an average of 30 unique mite species (almost certainly an overestimate; Walter and Proctor 2013), the impact on overall mite diversity would be limited relative to estimated mite diversity overall (e.g., 10,000 bird + 5000 mammal species × 30 mites per species = 450,000 mite species).

MITES ON PLANTS

Walter and Proctor (2013:285) stated that most woody dicots and many herbaceous dicots are attacked by eriophyoid mites (along with many monocots, conifers, and ferns). Furthermore, most eriophyoid mite species are host specific (e.g., approximately 80%; Skoracka et al. 2010), and a single plant species may be utilized by multiple mite species (e.g., sugar maple, *Acer saccharum*, attacked by eight mite species that are mostly host-specific; Patankar et al. 2012). Most of these species are thought to be undescribed (Amrine and Stasny 1994; Walter and Proctor 2013). Unfortunately, it is not clear what the typical number of mites per host plant species is (but presumably smaller plants have fewer mite species than a large tree such as a sugar maple). Assuming that there are ~340,000 plant species (see Methods), and as many as eight mite species per plant species, there could be as many as 2.7 million species of herbivorous mites. Furthermore, applying the number of cryptic arthropod species (six; see Appendix S1) would suggest up to 16.2 million mite species on plants. However, we emphasize that most plant species may have far fewer unique mite species, and that there might be only a few million eriophyoid species at most.

FREE-LIVING MITES

In addition to mites associated with other organisms, there is a vast diversity of free-living mites. Along with nematodes, mites are the most diverse organisms in soils around the world, and mites are more diverse at many sites, especially in temperate forests and some tropical forests (Wu et al. 2011). However, the global diversity of free-living mites is difficult to estimate. A single site near Hudson Bay (including multiple microhabitats, such as boreal forest) was estimated to have 1229 species of mites (approximately 900 identified based on DNA barcoding; Young et al. 2012). The turnover between sites within regions is unclear. Nevertheless, tropical mite richness might still exceed temperate richness at individual sites (Walter and Proctor 2013:Figure 11.2).

OVERALL MITE DIVERSITY

Estimates of overall mite diversity are particularly uncertain. We assume that there are ~6.8 million arthropod species projected based on morphological criteria (Stork et al. 2015), and that there are roughly six cryptic insect species for each morphological species (Appendix S1), and that these projections do not include insect-associated mites. This initial estimate yields 40.8 million arthropods. To then estimate the number of mites, we assumed that each of these arthropod species has (on average) one mite species uniquely associated with them, yielding 40.8 million additional arthropod species. We note that even if this is an overestimate of the number of insect-associated mites, it does not account for the many mite species associated with vertebrates and plants. Therefore, this estimate yields 81.6 million arthropods overall. We also performed analyses assuming reduced specificity of mites to their arthropod hosts, with a host:mite ratio of 1:0.25. This yields 10.2 million mite species, and 51 million arthropod species overall.

Phylum Nematoda

nematodes in insects

Nematodes are known to infect insects. Therefore, the proportional richness of nematodes (relative to other organisms) may depend on how many groups of insects are potentially infected by nematodes, the number of nematode parasites per insect species, and their level of host specificity. We know of few specific estimates for this. However, Powers et al. (2009) estimated that there are 0.8 unique endoparasitic nematode species per termite species in a Central American rainforest site, and summarized data from other studies suggesting that this ratio may be widespread for termite species in temperate and tropical sites. They also noted that each fig species appeared to contain two unique nematode species in their fruits, which appear to be related to fig wasps. They suggested that each plant and animal species at La Selva, Costa Rica might contain a unique nematode species. An important question is whether all insects contain similar specialized endoparasitic nematodes, and whether they have similar levels of host specificity.

The most diverse insect orders are Coleoptera, Diptera, Hymenoptera, and Lepidoptera. Grucmanová and Holuša (2013) summarized the nematode species found among one genus of beetles (*Ips*) in Central Europe. They found that each beetle species hosted an average 3.2 endoparasitic nematode species, of which 1.8 species are unique to each host, and 3.4 ectoparasitic nematode species, of which 1.4 are unique. Thus, there are an average of 3.2 unique nematode species per beetle species in this genus.

Diptera may host numerous nematodes. Simulid flies (for example) may host several genera of mermithid nematodes (Poinar 1977).

Camino and Achinelly (2011) described nematodes in two sympatric species of orthopterans in two different families, finding 16 and nine nematode species per host species, with 15 and eight unique nematode species per host species. However, since these are in different families, the level of host specificity at the species level is unclear. Nevertheless it is clear that orthopterans can harbor many nematode species, and that these can show considerable host specificity within the order.

In Hymenoptera, in addition to infecting wasps (e.g., Poinar 1977; Saito-Morooka 2014), nematodes are also known to infect ants (Poinar 2012). Many genera and families may infect a single host ant species or genus (the ant genus *Lasius* may host species in the nematode families Mermithidae, Diplogratridae, and Rhabditidae), and some families of nematodes have species that can be specific to a single ant species (e.g., Mermithidae; Poinar 2012).

Nematodes are known to infect Lepidoptera, but nematode ectoparasites appear to be relatively uncommon (Simmons and Rogers 1996). The prevalence of nematode endoparasites in Lepidoptera is unclear.

In summary, the most species-rich orders of insects (with the possible exception of Lepidoptera) seem to contain numerous nematode species, many of which appear to be host specific. Furthermore, those studies with relevant data suggest that the mean number of host-specific nematodes per insect host species may be at least one (including 0.8 for isopterans, 2 for fig wasps, and 3.2 for *Ips* beetles). We acknowledge that we have not shown that every major group of insects has host-specific nematodes. However, it seems difficult to explain why particular groups would be nematode-free, and we conservatively use an average of one host-specific nematode per host species, even though this might be higher.

nematodes parasitizing vertebrates

Poulin and Morand (2004) provided a summary of species richness and host specificity of nematodes parasitizing species of various major vertebrate clades. These allow one to calculate ratios of host to parasite richness for each group: Chondrichthyes: 0.180; Actinopterygii: 0.145; Amphibia: 0.535; Lepidosauria (reptiles): 1.014; Aves: 1.012; and Mammalia: 0.642. These are consistent with the idea that each terrestrial animal species might have roughly one unique nematode species, as suggested above for some insects (e.g., termites).

nematodes parasitizing plants

Nematodes are known to parasitize plants, including economically important pests that attack the roots of plants, such as the root-knot nematodes (*Meloidogyne* sp.) and cyst nematodes (*Heterodera* and *Globodera*; Williamson and Gleason 2003). The level of host specificity in these taxa can be quite wide. Given this, their global diversity patterns are uncertain, but might not be very high.

free-living nematodes

The species-richness of non-parasitic (free-living) nematodes is uncertain, but could be very high. The number of marine nematodes is particularly unclear. Some estimates suggested that

there were more than 1 million species, with this high richness driven largely by deep marine habitats, but these estimates have been challenged by the authors of some of these same estimates (Lambshead and Boucher 2003). However, some authors continue to cite Lambshead (2004), which suggested that total marine nematode richness might exceed 1 million species (of which only a few thousand are described). Importantly, the lower estimates from Lambshead and Boucher (2003) were actually made after the larger estimate from Lambshead (2004), even though their later estimate actually has an earlier publication date.

In addition to marine habitats, nematodes are the most diverse animals in the soil in some regions of the world (e.g., grasslands), along with mites (Wu et al. 2011). Using molecular barcodes and careful sampling, Powers et al. (2009) identified ~500 putative free-living species at a local site in the tropics (La Selva, Costa Rica). Most (approximately 66%) of the species were in the leaf litter and understory rather than in the soil. Powers et al. (2009) did not sample the canopy, however. Bloemers et al. (1997) identified 431 putative nematode species in the soil at a tropical rainforest reserve in Africa, of which 90% were undescribed. This location has the highest local soil nematode diversity (Powers et al. 2009).

The level of geographic turnover of the free-living nematode fauna remains unclear. Wu et al. (2011) found that very few soil nematode species were shared among sites in different regions (e.g., continents). A key question is how species composition of free-living nematodes at local sites changes over spatial scales within continents. Without this information we cannot speculate at global richness of free-living nematodes.

OVERALL NEMATODE DIVERSITY

We assume that the major driver of global nematode diversity is the set of nematodes species parasitizing insects. Given that there are at least 40.8 million free-living arthropods, there should be at least as many nematode species, as described above. An important but unresolved question is whether most mite species also tend to host species-specific nematodes. We initially assume that they do (Table 1), but also explored the impacts of assuming that mites host relatively few host-specific nematode species (Table 4). We also assume that nematodes parasitizing other animal phyla besides arthropods make a minor contribution to nematode diversity. For example, if there are only ~70,000 chordate species (Zhang 2013), with the majority of these being actinopterygian fish, and each has an average of only 0.145 unique nematode species, their nematode diversity may be quite limited. Other animal phyla also have nematodes, but their contribution may be relatively small as well. Similarly, the contribution of plant and soil-dwelling nematodes is unclear, but may be small relative to the 40.8 million suggested for those parasitizing insects. Therefore, we use a value of 40.8 million nematode species for our subsequent calculations.

OTHER ANIMAL PHYLA

Based on current numbers of described species, the most species rich animal phyla exclusive of Arthropoda and Nematoda are Mollusca (approximately 73,000 species), Chordata (about 70,000), Platyhelminthes (around 30,000), and Annelida (approximately 18,000; Zhang 2013). We addressed the diversity of Arthropoda and Nematoda above. There may be numerous undescribed species of these four orders. However, we are unaware of factors that would push the number of undescribed species in any of these groups into millions of species. For example, Chapman (2009) summarized estimates of projected global richness for many animal phyla, including Mollusca (around 200,000 species), Chordata (about 80,000), Platyhelminthes (approximately 80,000), and Annelida (around 30,000). In short, summing the estimates for these other animal phyla would add up to less than 0.5 million. Mora et al. (2011) estimated that there were 2.2 million animal species in the oceans, but did not focus on which phyla these species belonged to. Appeltans et al. (2012) suggested that Mora et al. (2011) had overestimated marine richness somewhat, and that there were only 0.7–1.0 million marine species (across all kingdoms), with the most diverse groups including Mollusca (about 150,000 estimated marine species), Arthropoda (approximately 150,000 crustaceans), Nematoda (61,400), Annelida (around 30,000), and Chromista (about 85,000). In summary, we argue that the contribution of these other animal phyla to global animal richness may be limited, especially in relation to our projected numbers for arthropods and nematodes.

TOTAL ANIMAL RICHNESS

As described above, we estimate that there are at least 81.6 million arthropod species, or potentially fewer if mites are less host specific in some insect groups (52 million). We estimate that there are 81.6 million nematode species, but possibly many fewer if mites host relatively few nematodes (40.8 million). Overall, we estimate that there 163.2 million animal species (but with estimates varying from 102 million).

REFERENCES

Amrine J. W. Jr., Stasny T. A. 1994. *Catalog of the Eriophyoidea (Acarina: Prostigmata) of the World*. West Boomfield (Michigan): Indira Publishing House.

Appeltans W., Ahyong S. T., Anderson G., et al. 2012. The magnitude of global marine species diversity. *Current Biology* 22:2189–2202.

Basset Y., Cizek L., Cuénoud P., et al. 2012. Arthropod diversity in a tropical forest. *Science* 338:1481–1484.

Beaulieu F., Dechene A. D., Walter D. E. 2008. Phase morphs and phoresy: new species of *Antennoseius* (*Vitzthumia*) mites (Acari: Mesostigmata: Ascidae) associated with pyrophilous carabids (Carabidae: *Sericoda* spp.) in Alberta, Canada. *Zootaxa* 1961: 37–57.

Bloemers G. F., Hodda M., Lambshead P. J. D., Lawton J. H., Wanless F. R. 1997. The effects of forest disturbance on diversity of tropical soil nematodes. *Oecologia* 111:575–582.

Camino N. B., Achinelly M. F. 2011. Biodiversity of insect-parasitic nematodes in soil pest insect (Orthoptera, Gryllidae and Gryllotalpidae) in wheat fields of Buenos Aires, Argentina. *Anales de Biología* 33:15–21.

Chapman A. D. 2009. *Numbers of Living Species in Australia and the World*. Second Edition. Canberra (Australia): Australian Biological Resources Study.

De Jong D., Morse R. A., Eickwort G. C. 1982. Mite pests of honey bees. *Annual Review of Entomology* 27:229–252.

Grucmanová Š., Holuša J. 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., Basset Y., Benke K. K., Grimbacher P. S., Miller S. E., Novotny V., Samuelson G. A., Stork N. E., Weiblen G. D., Yen J. D. L. 2010. Quantifying uncertainty in estimation of tropical arthropod species richness. *American Naturalist* 176:90–95.

Hamilton A. J., Basset Y., Benke K. K., Grimbacher P. S., Miller S. E., Novotny V., Samuelson G. A., Stork N. E., Weiben G. D., Yen J. D. L. 2011. Correction: quantifying uncertainty in estimation of tropical arthropod species richness. *American Naturalist* 177: 544–545.

Hamilton A. J., Novotny V., Waters E. K., Basset Y., Benke K. K., Grimbacher P. S., Miller S. E., Samuelson G. A., Weiblen G. D., Yen J. D. L., Stork N. E. 2013. Estimating global arthropod species richness: refining probabilistic models using probability bounds analysis. *Oecologia* 171:357–365.

Hunter P. E., Rosario R. M. T. 1988. Associations of Mesostigmata with other arthropods. *Annual Review of Entomology* 33:393–417.

Knee W., Beaulieu F., Skevington J. H., Kelso S., Cognato A. I., Forbes M. R. 2012. Species boundaries and host range of tortoise mites (Uropodoidea) phoretic on bark beetles (Scolytinae), using morphometric and molecular markers. *PLOS ONE* 7:e47243.

Knee W., Forbes M. R., Beaulieu F. 2013. Diversity and host use of mites (Acari: Mesostigmata, Oribatida) phoretic on bark beetles (Coleoptera: Scolytinae): global generalists, local specialists? *Annuals of the Entomological Society of America* 106:339–350.

Lambshead P. J. D. 2004. Marine nematode biodiversity. Pages 438–468 in *Nematology: Advances and Perspectives, Volume 1: Nematode Morphology, Physiology and Ecology*, edited by Z. X. Chen, S. Y. Chen, and D. W. Dickson. Wallingford (United Kingdom): CABI Publishing.

Lambshead P. J. D., Boucher G. 2003. Marine nematode deep-sea biodiversity—hyperdiverse or hype? *Journal of Biogeography* 30:475–485.

Mora C., Tittensor D. P., Adl S., Simpson A. G. B., Worm B. 2011. How many species are there on Earth and in the ocean? *PLOS Biology* 9:e1001127.

Patankar R., Beaulieu F., Smith S. M., Thomas S. C. 2012. The life history of a gall-inducing mite: summer phenology, predation and influence of gall morphology in a sugar maple canopy. *Agricultural and Forest Entomology* 14:251–259.

Poinar G. Jr. 2012. Nematode parasites and associates of ants: past and present. *Psyche* 2012:192017.

Poinar G. O. Jr. 1977. A synopsis of the nematodes occurring in blackflies (Diptera: Simuliidae). *Bulletin of the World Health Organization* 55:509–515.

Poulin R., Morand S. 2004. *Parasite Biodiversity*. Washington (DC): Smithsonian Books.

Powers T. O., Neher D. A., Mullin P., Esquivel A., Giblin-Davis R. M., Kanzaki N., Stock S. P., Mora M. M., Uribe-Lorio L. 2009. Tropical nematode diversity: vertical stratification of nematode communities in a Costa Rican humid lowland rainforest. *Molecular Ecology* 18:985–996.

Roskov Y. et al. 2014. Species 2000 and ITIS Catalogue of Life, 29 October 2014. Leiden (The Netherlands): Science 2000, 2004. Available at http://www.catalogueoflife.org/col.

Saito-Morooka F. 2014. The prevalence of the parasitic nematode *Sphaerularia* sp. in the overwintering gynes of *Parapolybia* spp. (Hymenoptera, Polistinae). *Journal of Hymenopteran Research* 38:37–43.

Salmane I., Telnov D. 2009. Mesostigmata mites (Acari: Parasitiformes) associated with beetles (Insecta: Coleoptera) in Latvia. *Latvijas Entomologs* 47:58–70.

Schwarz H. H., Starrach M., Koulianos S. 1998. Host specificity and permanence of associations between mesostigmatic mites (Acari: Anactinotrichida) and burying beetles (Coleoptera: Silphidae: *Nicrophorus*). *Journal of Natural History* 32:159–172.

Simmons A. M., Rogers C. E. 1996. Ectoparasitic acugutturid nematodes of adult Lepidoptera. *Journal of Nematology* 28:1–7.

Skoracka A., Smith L., Oldfield G., Cristofaro M., Amrine J. W. 2010. Host-plant specificity and specialization in eriophyoid mites and their importance for the use of eriophyoid mites as biocontrol agents of weeds. *Experimental and Applied Acarology* 51:93–113.

Stork N. E., McBroom J., Gely C., Hamilton A. J. 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.

Walter D. E., Proctor H. C. 2013. *Mites: Ecology, Evolution, and Behaviour: Life at a Microscale*. Second Edition. Dordrecht (The Netherlands): Springer.

Williamson V. M., Gleason C. A. 2003. Plant-nematode interactions. *Current Opinion in Plant Biology* 6:327–333.

Wu T., Ayres E., Bardgett R. D., Wall D. H., Garey J. R. 2011. Molecular study of worldwide distribution and diversity of soil animals. *Proceedings of the National Academy of Sciences of the United States of America* 108:17720–17725.

Young M. R., Behan-Pelletier V. M., Hebert P. D. N. 2012. Revealing the hyperdiverse mite fauna of subarctic Canada through DNA barcoding. *PLOS ONE* 7:e48755.

Zhang Z. 2013. Animal biodiversity: an update of classification and diversity in 2013. *Zootaxa* 3703:005–011.

APPENDIX S3

Estimating Fungal Richness

Overview

Fungi are among the most ecologically important organisms on Earth (Mueller and Schmit 2007) and the extent of fungal diversity has been debated for decades. Although there are only ~128,432 currently described species (Roskov et al. 2014), many researchers have suggested that global diversity is actually much higher. Many estimates have been based on extrapolations using the ratio of fungal species to plant species in well-studied areas. For example, Hawksworth (1991) estimated 1.62 million fungal species based upon a 6:1 ratio of fungal to plant species in Great Britain. This approach has been criticized by some on the basis of its generalization from small to very large scales (May 1991, 1994; Mueller and Schmit 2007; Schmit and Mueller 2007). However, Hawksworth's (1991) estimate has nonetheless become widely accepted. Furthermore, other estimates of global fungal diversity have approached 10 million species (Bass and Richards 2011 and references therein). Bass and Richards (2011) concluded that there is minimally one order of magnitude more fungal species than the ~128,000 currently known.

In the sections below, we briefly characterize some groups of fungi that may harbor vast undescribed species richness. We focus specifically on groups that have sufficient evidence for generalization about species numbers and that have important implications for the total number of fungal species, especially those associated with animals. We emphasize that many groups of fungi are not addressed in our analysis, including lichens, plant pathogens, mycorrhizal fungi, and arthropod gut fungi (see Blackwell 2011 for other potentially important reservoirs of fungal diversity). These groups were excluded either because we considered them unlikely to have major impacts on our estimates or because the existing literature did not permit an estimate of their overall species richness. For example, mycorrhizal fungi exhibit both low species richness and low host specificity (Lee et al. 2013; van der Heijden et al. 2015) and are therefore excluded from our analysis. Similarly, arthropod gut fungi (class Trichomycetes) may only exhibit host specificity at the level of host families or genera (Lichtwardt et al. 2001; Cafaro 2002).

Soil Fungi

updated estimate from fungus:plant ratios

Building on the fungal to plant ratio (F:P) approach, Taylor et al. (2014) used direct molecular techniques to comprehensively census soil fungi in a boreal forest ecosystem with well-known plant diversity. They found a regionally consistent F:P ratio of at least 17:1. Assuming 352,000 vascular plant species globally, Taylor et al. (2014) extrapolated to a total of ~6 million soil fungal species. However, Tedersoo et al. (2014) showed that F:P ratios increase dramatically from tropical to polar latitudes, and therefore that extrapolations of fungal richness based on F:P ratios in boreal latitudes overestimate soil fungal richness by ~2.5 fold. Taking into account the estimate from Taylor et al. (2014) as well as variation in F:P ratios discovered by Tedersoo et al. (2014), we estimate approximately 2.4 million soil fungal species. Importantly, Taylor et al. (2014) noted that their estimate was conservative for a number of reasons, most importantly in that their method of species delimitation tends to lump together sequences from closely related but distinct taxa. As such, a

global estimate of 2.4 million soil fungi may be a lower boundary.

Animal-Associated Fungi

microsporidia

The microsporidia are a group of obligate intracellular fungal parasites that have been documented across most animal phyla (Keeling and Slamovits 2004), including arthropods (see below) and nematodes (Troemel et al. 2008; Ardila-Garcia and Fast 2012). They are particularly well studied in arthropods and fish (Keeling and Fast 2002; Smith 2009). Notably, microsporidia have been documented in the most diverse insect orders, including Coleoptera (e.g., Yaman et al. 2010; Kyei-Poku et al. 2011; Ovcharenko et al. 2013), Lepidoptera (e.g., Solter et al. 2000), and Hymenoptera and Diptera (see references below). Some microsporidian species can infect many distantly related hosts, but specificity to a single host or a group of closely related hosts may predominate (Baker et al. 1998; Keeling and Fast 2002; Smith 2009; Vávra and Lukeš 2013). For example, although some microsporidia can infect hosts from multiple lepidopteran families in laboratory settings, ecological host specificity is generally much narrower than physiological host specificity (Solter and Maddox 1998; Solter et al. 2000). Subsequent field experiments have confirmed narrow host specificity of microsporidia associated with Lepidoptera (Solter et al. 2010). Further, it has been suggested that microsporidia have frequently cospeciated with their hosts (Vávra and Lukeš 2013). For example, Shafer et al. (2009) found evidence of both strict-sense cospeciation and host switching between four species of *Nosema* and their bee hosts. Similarly, Andreadis et al. (2012) found compelling evidence of joint speciation between *Amblyospora* species and *Culex* mosquitoes, with one microsporidian species per host species. Moreover, although some host-parasite associations indicated host switching, there were several instances where single mosquito species hosted two to four microsporidian species each, suggesting that microsporidians may also diversify independently in isolated populations of their hosts (Andreadis et al. 2012). Cospeciation has also been indicated by strong congruence between the phylogenies of *Loma* species and their fish hosts in the Pacific (Brown et al. 2010). Given these findings, the hypothesis that microsporidian richness may equal animal richness (Keeling and Fast 2002) may actually be conservative. We therefore assume that each animal species hosts at least one unique microsporidian species.

entomopathogenic fungi

Entomopathogenic fungi have been extensively studied, given their potential use for biological control of insect pests (Vega et al. 2012). Most research has focused on Entomophthorales (Zygomycota) and Hypocreales (Ascomycota), which occur in most insect orders (Keller 2007). Although the degree of host specialization varies in both groups, species in Hypocreales are considered facultative host generalists whereas those in Entomophthorales are considered highly host-specific (Roy et al. 2006; Vega et al. 2012; Boomsma et al. 2014). For example, species-level host specificity has been documented in fly-pathogenic *Entomophthora* (Jensen et al. 2001, 2006). However, it is unclear if this degree of host specificity is mirrored in *Entomophthora* associated with other insect orders. For example, host-driven divergence of *Entomophthora* appears to be much less extensive in aphids (Jensen et al. 2009). Similarly, despite the generalization that species of Hypocreales have broad host ranges, some lineages appear to be highly host-specific. For example, members of the genus *Ophiocordyceps* have been shown to be specific to their ant hosts (Evans et al. 2011; Kobmoo et al. 2012), and some recently discovered *Ophiocordyceps* species also appear to have strict host affiliations (Kobmoo et al. 2015; Sanjuan et al. 2015). Overall, it appears that at least some insect lineages harbor species-specific entomophathogenic fungi. These entomopathogenic fungi may be as diverse as Microsporidia, but they also exhibit host generalism. Given the latter pattern, we do not include them in our calculations, but we acknowledge that they might be as diverse as microsporidians appear to be.

Plant-Associated Fungi

fungal endophytes

Fungal endophytes occur in all major plant lineages, and occur from tropical to arctic habitats (Arnold et al. 2000; Arnold 2007). Endophyte species richness has been documented in many studies, which frequently reveal ~10 to 50 fungal species in a given host plant species (e.g., Arnold and Lutzoni 2007), but with the number of fungal species unique to each host plant species remaining uncertain. Even higher species richness has been documented in studies with extensive sampling within a single host species, for example, more than 4000 operational taxonomic units (OTUs) in a tree species endemic to the Hawaiian Islands (Zimmerman and Vitousek 2012). In these studies, putative species (OTUs) were delineated

based on DNA sequence similarity. Despite suggestions that fungal endophytes may be hyperdiverse (Arnold et al. 2000), uncertainty surrounding species detection and host specificity makes it difficult to estimate their overall species richness. For example, studies of endophyte diversity are sometimes limited by non-asymptotic species accumulation curves (e.g., Arnold and Lutzoni 2007; Zimmerman and Vitousek 2012), suggesting incomplete sampling. Furthermore, large proportions of detected endophytic OTUs are singletons (e.g., Arnold and Lutzoni 2007; Hoffman and Arnold 2008; Lau et al. 2013; Higgins et al. 2014; Massimo et al. 2015), meaning that they are known from a single sampled individual. This is problematic because it may be difficult to distinguish rare individuals from those that are truly specific to a given host species (U'Ren 2011). Even when analyses are restricted to non-singleton taxa, patterns in host specificity are highly variable, with some studies finding narrow host affiliations (e.g., Higgins et al. 2007; Massimo et al. 2015) and others find striking host generalism (e.g., Higgins et al. 2007, 2011; Sandberg et al. 2014). Therefore, we do not include endophytes in our overall estimate of fungal richness. In the future, targeted studies of closely related pairs of host taxa with extensive sampling within each host species should help clarify the extent of endophyte species richness and host specificity. However, we note that even if there were (on average) ~50 fungal species unique to each plant species, the impact of endophytes on overall fungal richness would still be relatively limited (e.g., approximately 18 million added to an estimated total of about 163 million; Table 1).

REFERENCES

Andreadis T. G., Simakova A. V., Vossbrinck C. R., Shepard J. J., Yurchenko Y. A. 2012. Ultrastructural characterization and comparative phylogenetic analysis of new microsporidia from Siberian mosquitoes: evidence for coevolution and host switching. *Journal of Invertebrate Pathology* 109:59–75.

Ardila-Garcia A.-M., Fast N. M. 2012. Microsporidian infection in a free-living marine nematode. *Eukaryotic Cell* 11:1544–1551.

Arnold A. E. 2007. Understanding the diversity of foliar endophytic fungi: progress, challenges, and frontiers. *Fungal Biology Reviews* 21:51–66.

Arnold A. E., Lutzoni F. 2007. Diversity and host range of foliar fungal endophytes: are tropical leaves biodiversity hotspots? *Ecology* 88:541–549.

Arnold A. E., Maynard Z., Gilbert G. S., Coley P. D., Kursar T. A. 2000. Are tropical fungal endophytes hyperdiverse? *Ecology Letters* 3:267–274.

Baker M. D., Vossbrinck C. R., Becnel J. J., Andreadis T. G. 1998. Phylogeny of *Amblyospora* (Microsporidia: Amblyosporidae) and related genera based on small subunit ribosomal DNA data: a possible example of host parasite cospeciation. *Journal of Invertebrate Pathology* 71:199–206.

Bass D., Richards T. A. 2011. Three reasons to re-evaluate fungal diversity "on Earth and in the ocean." *Fungal Biology Reviews* 25:159–164.

Blackwell M. 2011. The fungi: 1, 2, 3 . . . 5.1 million species? *American Journal of Botany* 98:426–438.

Boomsma J. J., Jensen A. B., Meyling N. V., Eilenberg J. 2014. Evolutionary interaction networks of insect pathogenic fungi. *Annual Review of Entomology* 59: 467–485.

Brown A. M. V., Kent M. L., Adamson M. L. 2010. Description of five new *Loma* (Microsporidia) species in Pacific fishes with redesignation of the type species *Loma morhua* Morrison & Sprague, 1981, based on morphological and molecular species-boundaries tests. *Journal of Eukaryotic Microbiology* 57:529–553.

Cafaro M. J. 2002. Species richness patterns in symbiotic gut fungi (Trichomycetes). *Fungal Diversity* 9: 47–56.

Evans H. C., Elliot S. L., Hughes D. P. 2011. Hidden diversity behind the zombie-ant fungus *Ophiocordyceps unilateralis*: four new species described from carpenter ants in Minas Gerais, Brazil. *PLOS ONE* 6:e17024.

Hawksworth D. L. 1991. The fungal dimension of biodiversity: magnitude, significance, and conservation. *Mycological Research* 95:641–655.

Higgins K. L., Arnold A. E., Miadlikowska J., Sarvate S. D., Lutzoni F. 2007. Phylogenetic relationships, host affinity, and geographic structure of boreal and arctic endophytes from three major plant lineages. *Molecular Phylogenetics and Evolution* 42:543–555.

Higgins K. L., Coley P. D., Kursar T. A., Arnold A. E. 2011. Culturing and direct PCR suggest prevalent host generalism among diverse fungal endophytes of tropical grasses. *Mycologia* 103:247–260.

Higgins K. L., Arnold A. E., Coley P. D., Kursar T. A. 2014. Communities of fungal endophytes in tropical forest grasses: highly diverse host- and habitat generalists characterized by strong spatial structure. *Fungal Ecology* 8:1–11.

Hoffman M. T., Arnold A. E. 2008. Geographic locality and host identity shape fungal endophyte commu-

nities in cupressaceous trees. *Mycological Research* 112:331–344.

Jensen A. B., Thomsen L., Eilenberg J. 2001. Intraspecific variation and host specificity of *Entomophthora muscae sensu stricto* isolates revealed by random amplified polymorphic DNA, universal primed PCR, PCR-restriction fragment length polymorphism, and conidial morphology. *Journal of Invertebrate Pathology* 78:251–259.

Jensen A. B., Thomsen L., Eilenberg J. 2006. Value of host range, morphological, and genetic characteristics within the *Entomophthora muscae* species complex. *Mycological Research* 110:941–950.

Jensen A. B., Eilenberg J., Lastra C. L. 2009. Differential divergences of obligately insect-pathogenic *Entomophthora* species from fly and aphid hosts. *FEMS Microbiology Letters* 300:180–187.

Keeling P. J., Fast N. M. 2002. Microsporidia: biology and evolution of highly reduced intracellular parasites. *Annual Review of Microbiology* 56:93–116.

Keeling P. J., Slamovits C. H. 2004. Simplicity and complexity of microsporidian genomes. *Eukaryotic Cell* 3:1363–1369.

Keller S. 2007. *Arthropod-Pathogenic Entomophthorales: Biology, Ecology, Identification*. COST Action 842. Luxembourg: Office for Official Publications of the European Communities.

Kobmoo N., Mongkolsamrit S., Tasanathai K., Thanakitpipattana D., Luangsa-ard J. J. 2012. Molecular phylogenies reveal host-specific divergence of *Ophiocordyceps unilateralis sensu lato* following its host ants. *Molecular Ecology* 21:3022–3031.

Kobmoo N., Mongkolsamrit S., Wutikhun T., Tasanathai K., Khonsanit A., Thanakitpipattana D., Luangsa-ard J. J. 2015. New species of *Ophiocordyceps unilateralis*, an ubiquitous pathogen of ants from Thailand. *Fungal Biology* 119:44–52.

Kyei-Poku G., Gauthier D., Schwarz R., van Frankenhuyzen K. 2011. Morphology, molecular characteristics and prevalence of a *Cystosporogenes* species (Microsporidia) isolated from *Agrilus anxius* (Coleoptera: Buprestidae). *Journal of Invertebrate Pathology* 107:1–10.

Lau M. K., Arnold A. E., Johnson N. C. 2013. Factors influencing communities of foliar fungal endophytes in riparian woody plants. *Fungal Ecology* 6: 365–378.

Lee E.-H., Eo J.-K., Ka K.-H., Eom A.-H. 2013. Diversity of arbuscular mycorrhizal fungi and their roles in ecosystems. *Mycobiology* 41:121–125.

Lichtwardt R. W., Cafaro M. J., White M. M. 2001. The Trichomycetes: Fungal Associates of Arthropods. Revised Edition. Available at http://www.nhm.ku.edu/~fungi/Monograph/Text/Mono.htm.

Massimo N. C., Nandi Devan M. M., Arendt K. R., Wilch M. H., Riddle J. M., Furr S. H., Steen C., U'Ren J. M., Sandberg D. C., Arnold A. E. 2015. Fungal endophytes in aboveground tissues of desert plants: infrequent in culture, but highly diverse and distinctive symbionts. *Microbial Ecology* 70:61–76.

May R. M. 1991. A fondness for fungi. *Nature* 352:475–476.

May R. M. 1994. Conceptual aspects of quantification of the extent of biological diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences* 345:13–20.

Mueller G. M., Schmit J. P. 2007. Fungal biodiversity: what do we know? What can we predict? *Biodiversity and Conservation* 16:1–5.

Ovcharenko M., Świątek P., Ironside J., Skalski T. 2013. *Orthosomella lipae* sp. n. (Microsporidia) a parasite of the weevil, *Liophloeus lentus* Germar, 1824 (Coleoptera: Curculionidae). *Journal of Invertebrate Pathology* 112:33–40.

Roskov Y. et al. 2014. Species 2000 and ITIS Catalogue of Life, 29 October 2014. Leiden (The Netherlands): Species 2000, 2014. Available at http://www.catalogueoflife.org/col.

Roy H. E., Steinkraus D. C., Eilenberg J., Hajek A. E., Pell J. K. 2006. Bizarre interactions and endgames: entomopathogenic fungi and their arthropod hosts. *Annual Review of Entomology* 51:331–357.

Sandberg D. C., Battista L. J., Arnold A. E. 2014. Fungal endophytes of aquatic macrophytes: diverse host-generalists characterized by tissue preferences and geographic structure. *Microbial Ecology* 67:735–747.

Sanjuan T. I., Franco-Molano A. E., Kepler R. M., Spatafora J. W., Tabima J., Vasco-Palacios A. M., Restrepo S. 2015. Five new species of entomopathogenic fungi from the Amazon and evolution of neotropical *Ophiocordyceps*. *Fungal Biology* 119:901–916.

Schmit J. P., Mueller G. M. 2007. An estimate of the lower limit of global fungal diversity. *Biodiversity and Conservation* 16:99–111.

Shafer A. B. A., Williams G. R., Shutler D., Rodgers R. E. L., Stewart D. T. 2009. Cophylogeny of *Nosema* (Microsporidia: Nosematidae) and bees (Hymenoptera: Apidae) suggests both cospeciation and a host-switch. *Journal of Parasitology* 95:198–203.

Smith J. E. 2009. The ecology and evolution of microsporidian parasites. *Parasitology* 136:1901–1914.

Solter L. F., Maddox J. V. 1998. Physiological host specificity of Microsporidia as an indicator of ecological host specificity. *Journal of Invertebrate Pathology* 71:207–216.

Solter L. F., Pilarska D. K., Vossbrinck C. F. 2000. Host specificity of Microsporidia pathogenic to forest Lepidoptera. *Biological Control* 19:48–56.

Solter L. F., Pilarska D. K., McManus M. L., Zúbrik M., Patočka J., Huang W.-F., Novotný J. 2010. Host specificity of microsporidia pathogenic to the gypsy moth, *Lymantria dispar* (L.): field studies in Slovakia. *Journal of Invertebrate Pathology* 105:1–10.

Taylor D. L., Hollingsworth T. N., McFarland J. W., Lennon N. J., Nusbaum C., Ruess R. W. 2014. A

first comprehensive census of fungi in soil reveals both hyperdiversity and fine-scale niche partitioning. *Ecological Monographs* 84:3–20.

Tedersoo L., Bahram M., Põlme S., et al. 2014. Global diversity and geography of soil fungi. *Science* 346: 1256688.

Troemel E. R., Félix M.-A., Whiteman N. K., Barrière A., Ausubel F. M. 2008. Microsporidia are natural intracellular parasites of the nematode *Caenorhabditis elegans*. *PLOS Biology* 6:e309.

U'Ren J. M. Host-, geographic- and ecological specificity of endophytic and endolichenic fungal communities. PhD diss., University of Arizona, 2011.

van der Heijden M. G. A., Martin F. M., Selosse M.-A., Sanders I. R. 2015. Mycorrhizal ecology and evolution: the past, the present, and the future. *New Phytologist* 205:1406–1423.

Vávra J., Lukeš J. 2013. Microsporidia and "the art of living together." *Advances in Parasitology* 82:253–319.

Vega F. E., Meyling N. V., Luangsa-ard J. J., Blackwell M. 2012. Fungal entomopathogens. Pages 171–220 in *Insect Pathology*, Second Edition, edited by F. E. Vega and H. K. Kaya. London (United Kingdom): Academic Press.

Yaman M., Radek R., Weiser J., Toguebaye B. S. 2010. *Unikaryon phyllotretae* sp. n. (Protista, Microspora), a new microsporidian pathogen of *Phyllotreta undulata* (Coleoptera; Chrysomelidae). *European Journal of Protistology* 46:10–16.

Zimmerman N. B., Vitousek P. M. 2012. Fungal endophyte communities reflect environmental structuring across a Hawaiian landscape. *Proceedings of the National Academy of Sciences of the United States of America* 109:13022–13027.

APPENDIX S4

Estimating Protist Richness

Protists are a polyphyletic assemblage of eukaryotic clades that are very poorly understood with respect to undescribed species richness (Pawlowski et al. 2012). Many researchers historically believed that protists had low global richness due to the ubiquity of a few cosmopolitan morphospecies (Finlay and Fenchel 1999; Finlay 2004). However, this view is in conflict with the influx of studies reporting high sequence diversity in environmental samples across habitats, with many novel sequences representing new species (Foissner 1999) and possibly new higher-level clades (Moon-van der Staay et al. 2001; Šlapeta et al. 2005; Epstein and López-García 2008; Massana and Pedrós-Alió 2008; de Vargas et al. 2015).

Estimates of actual protist richness are often several times higher than the number of described species. Mora et al. (2011) predicted 27,500 species based on extrapolation of 13,033 cataloged species, but cautioned that the polyphyly and taxonomic instability of protist lineages complicated this estimation. Appeltans et al. (2012) predicted marine richness alone to be 77,930–93,923 species within Chromista and 2207 within Protozoa (compared to 19,444 and 542 cataloged species, respectively, based on expert opinion). Recently, de Vargas et al. (2015) estimated 150,000 putative species of planktonic eukaryotes in the marine photic zone globally, of which 85% were protists. Pawlowski et al. (2012) updated the number of cataloged morphospecies from 26,010 to 74,373. They predicted global richness to be between 140,000 and 1.6 million species based on unknown sequences recovered in environmental DNA samples reported in Adl et al. (2007), a study that also predicted 1.2 million species of parasitic apicomplexans.

As with animals and fungi, the greatest potential for large-scale undescribed protist richness may come from parasitic groups (Roberts and Janovy 2009). It has been suggested that all animal species have at least one host-specific species of Apicomplexa (Adl et al. 2007; Cotterill et al. 2009; Morrison 2009; Pawlowski et al. 2012). Apicomplexans include the famous parasites *Toxoplasmosis* and *Plasmodium* (i.e., malaria). Apicomplexans infect most (and possibly all) major invertebrate taxa (Sparks 1985) such as free-living nematodes (Poinar and Hess 1988; information is lacking for endoparasitic nematodes) and all arthropod lineages (Levine 1988; Golemansky and Lipa 1991; Tanada and Kaya 1993), including beetles (Matthes and Guhl 1975). In addition, host specificity is often found to be strict when thoroughly investigated among closely related host species (birds: Bensch et al. 2000; Beadell et al. 2004; Blattodea: Clopton and Gold 1996; Smith and Cook 2008; Diptera: Lantová et al. 2010; Coleoptera and other insect orders: Clopton 2009). Apicomplexan parasite species may also be specific to different life stages of the same host species (as in mealworm beetles; Clopton et al. 1992). If this latter phenomenon is widespread in holometabolous insects, there is potential for overall species richness of apicomplexans to be

extraordinarily high. However, this pattern has not yet been widely established across arthropod hosts.

In the main text, we assume one species of parasitic protist species per animal host species as a reasonable and conservative estimate congruent with predictions from other authors (Adl et al. 2007; Cotterill et al. 2009; Morrison 2009). In contrast, although much higher than previously anticipated, free-living protist diversity may not form a large proportion of overall protist richness compared to symbionts.

REFERENCES

Adl S. M., Leander B. S., Simpson A. G. B., Archibald J. M., Anderson O. R., Bass D., Bowser S. S., Brugerolle G., Farmer M. A., Karpov S., Kolisko M., Lane C. E., Lodge D. J., Mann D. G., Meisterfeld R., Mendoza L., Moestrup Ø., Mozley-Standridge S. E., Smirnov A. V., Spiegel F. 2007. Diversity, nomenclature, and taxonomy of protists. *Systematic Biology* 56: 684–689.

Appeltans W., Ahyong S. T., Anderson G., et al. 2012. The magnitude of global marine species diversity. *Current Biology* 22:2189–2202.

Beadell J. S., Gering E., Austin J., Dumbacher J. P., Peirce M. A., Pratt T. K., Atkinson C. T., Fleischer R. C. 2004. Prevalence and differential host-specificity of two avian blood parasite genera in the Australo-Papuan region. *Molecular Ecology* 13:3829–3844.

Bensch S., Stjernman M., Hasselquist D., Örjan Ö., Hannson B., Westerdahl H., Pinheiro R. T. 2000. Host specificity in avian blood parasites: a study of *Plasmodium* and *Haemoproteus* mitochondrial DNA amplified from birds. *Proceedings of the Royal Society B: Biological Sciences* 267:1583–1589.

Clopton R. E. 2009. Phylogenetic relationships, evolution, and systematic revision of the septate gregarines (Apicomplexa: Eugregarinorida: *Septatorina*). *Comparative Parasitology* 76:167–190.

Clopton R. E., Gold R. E. 1996. Host specificity of *Gregarina blattarumvon* Siebold, 1839 (Apicomplexa: Eugregarinida) among five species of domiciliary cockroaches. *Journal of Invertebrate Pathology* 67:219–223.

Clopton R. E., Janovy J. Jr., Percival T. J. 1992. Host stadium specificity in the gregarine assemblage parasitizing *Tenebrio molitor*. *Journal of Parasitology* 78: 334–337.

Cotterill F. P. D., Al-Rasheid K., Foissner W. 2009. Conservation of protists: is it needed at all? Pages 193–209 in *Protist Diversity and Geographical Distribution*, edited by W. Foissner and D. L. Hawksworth. Heidelberg (Germany): Springer Science.

de Vargas C., Audic S., Henry N., et al. 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science* 348:1261605.

Epstein S., López-García P. 2008. "Missing" protists: a molecular prospective. *Biodiversity and Conservation* 17:261–276.

Finlay B. J. 2004. Protist taxonomy: an ecological perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences* 359:599–610.

Finlay B. J., Fenchel T. 1999. Divergent perspectives on protist species richness. *Protist* 150:229–233.

Foissner W. 1999. Protist diversity: estimates of the near-imponderable. *Protist* 150:363–368.

Golemansky V., Lipa J. J. 1991. Gregarines (Apicomplexa: Eugregarinida) from cave and terrestrial arthropods in Bulgaria. *Acta Protozoologica* 30:201–208.

Lantová L., Ghosh K., Svobodová M., Braig H. R., Rowton E., Weina P., Volf P., Votýpka J. 2010. The life cycle and host specificity of *Psychodiella sergenti* n. sp. and *Ps. tobbi* n. sp. (Protozoa: Apicomplexa) in sand flies *Phlebotomus sergenti* and *Ph. tobbi* (Diptera: Psychodidae). *Journal of Invertebrate Pathology* 105:182–189.

Levine N. D. 1988. *The Protozoan Phylum Apicomplexa*, Volume I. Boca Raton (Florida): CRC Press.

Massana R., Pedrós-Alió C. 2008. Unveiling new microbial eukaryotes in the surface ocean. *Current Opinion in Microbiology* 11:213–218.

Matthes D., Guhl W. 1975. Systematik, Anpassungen und Raumparasitismus auf Hydrophiliden lebender opercullariformer Epistyliden. *Arch Protistenk* 117: 110–186.

Moon-van der Staay S. Y., De Wachter R., Vaulot D. 2001. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* 409: 607–610.

Mora C., Tittensor D. P., Adl S., Simpson A. G. B., Worm B. 2011. How many species are there on Earth and in the ocean? *PLOS Biology* 9:e1001127.

Morrison D. A. 2009. Evolution of the Apicomplexa: where are we now? *Trends in Parasitology* 25:375–382.

Pawlowski J., Audic S., Adl S., et al. 2012. CBOL Protist Working Group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLOS Biology* 10:e1001419.

Poinar G. O. Jr., Hess R. 1988. Protozoan diseases. Pages 103–131 in *Diseases of Nematodes*, Volume 1, edited by G. O. Poinar Jr. and H. B. Jansson. Boca Raton (Florida): CRC Press.

Roberts L. S., Janovy J. 2009. *Gerald D. Schmidt and Larry S. Roberts' Foundations of Parasitology*. Eighth Edition. Boston (Massachusetts): McGraw-Hill Higher Education.

Šlapeta J., Moreira D., López-García P. 2005. The extent of protist diversity: insights from molecular ecology of freshwater eukaryotes. *Proceedings of the Royal Society B: Biological Sciences* 272:2073–2081.

Smith A. J., Cook T. J. 2008. Host specificity of five species of *Eugregarinida* among six species of cockroaches (Insecta: Blattodea). *Comparative Parasitology* 75:288–291.

Sparks A. K. 1985. *Synopsis of Invertebrate Pathology: Exclusive of Insects*. Amsterdam (The Netherlands): Elsevier Science Publishers.

Tanada Y., Kaya H. K. 1993. *Insect Pathology*. San Diego (California): Academic Press.

APPENDIX S5

Estimating Bacterial Richness

Overview

In order to estimate global species richness, it is important to include bacteria, the most abundant living organisms on the planet (Whitman et al. 1998). Less than 1% of all bacteria can be cultured, which historically has limited the detection of their diversity (Torsvik et al. 1990; Amann et al. 1995; Pace 1997). Therefore, new molecular techniques that can generate millions of sequences in a single sequencing run allow a much deeper analysis of bacterial diversity. We suggest that bacteria that are associated with animals are more diverse than previously recognized and may be the numerically dominant organisms on the planet. Free-living bacteria may also be more diverse than currently recognized, however, the richness of bacterial species associated with animals may be even greater.

Bacteria Associated with Animals

Excellent reviews and meta-analyses of bacterial diversity associated with animal hosts already exist and were used extensively here. Some of these reviews are general and describe patterns of bacterial diversity across different animal hosts and environments (Ley et al. 2008; Nemergut et al. 2011), while others are specific to certain host groups: mammals (Ley et al. 2008), fish (Sullam et al. 2012), birds (Waite et al. 2014), soil nematodes (Ladygina et al. 2009), mites (Chaisiri et al. 2015), insects (Colman et al. 2012; Engel and Moran 2013; Jones et al. 2013; Yun et al. 2014), and marine sponges (Schmitt et al. 2012). These representative groups and their associated bacterial diversity are shown in Table S3. This table is not meant to be exhaustive, but merely to demonstrate the diversity of the associated bacteria that are present across animal phyla.

Table S3 lists the average number of bacterial operational taxonomic units (OTUs) for a host species separated by taxonomic order. In molecular studies of bacteria, sequences are typically clustered together that are less than 3% different at the DNA level and described as OTUs. OTUs are used as a proxy for "species" in these molecular assessments. The lowest number of bacterial OTUs within a host species across animal orders examined is Hemiptera with three. The highest bacterial diversity per host species is within a single species of mites (Sarcoptiformes) with 557 unique OTUs, however, we note this estimate is much larger than other mite-associated bacterial species estimates and may be an overestimate. Taking an average of these data provides a rough estimate of 83 OTUs of bacteria per animal host species. However, given the limited amount of host specificity data associated with these studies, we cannot use this number other than as a rough estimate of the bacterial diversity that is associated with animals.

Host specificity and calculations of unique bacterial species per host animal species

Data from previously published research were used to estimate insect gut-associated microbial diversity. Only studies that contained estimates of species richness within a single host insect genus were used in order to estimate the number of host-specific bacterial species. Using these data, the number of bacterial species per insect host can be estimated.

Chandler et al. (2011) examined the number of OTUs that were unique between species of *Drosophila* (Diptera) and two species from closely related genera sampled from a variety of locations in Asia, Africa, and North America (two unidentified *Drosophila* species, *D. elegans, D. flavohirta, D. falleni, D. hydei, D. immigrans, D. sulfurigaster, D. melanogaster, D. mojavensis* + *D. arizonae, D. sechellia, D. takahashii, Microdrosophila* sp., *Scaptodrosophila hibiscii*). A total of 7–20 individuals were pooled

from each sampling locale and species. Universal 16S primers were used to amplify the gene and were then cloned and sequenced with Sanger sequencing. The mean number of OTUs for each of these species was 15. Furthermore, on average, 89.3% of the OTUs were unique to a species in each pairwise comparison. Therefore, we estimate that there are 13.4 bacterial species unique to each *Drosophila* host species.

Sanders et al. (2014) estimated the number of bacterial OTUs across species from the genus *Cephalotes* (turtle ants; Hymenoptera). The genus *Cephalotes* contains around 130 described species, however, in our calculations we only focused on three clades that included closely related species to better capture the number of unique bacterial OTUs per host species. Data were obtained from their Table S1 and from the datadryad link listed in the paper. In the *laminatus* clade, the four sampled species (*C. minutus, C. pusillus, C. simillimus, C. spinosus*) were used to calculate the average number of OTUs and the average number of OTUs unique to each host species. Samples were collected in Peru or Brazil and a total of three individuals were pooled for each sample. These four species averaged 19 bacterial OTUs per species, and on average only nine were shared in each species comparison (52.7% unique; 10 unique bacterial species per OTU). For the *depressus* clade, three species were collected, again either in Brazil or Peru, and a total of three individuals in each case (*C. borgmeieri, C. cordatus, C. eduarduli*). In this case the average number of OTUs was 16, with an average number of eight that were shared (eight unique bacterial species per host species). These two estimates of unique bacterial species per host species were averaged to obtain a mean of nine bacterial species per host species for the genus *Cephalotes*.

Finally, data were obtained for three species of wasps (Hymenoptera) from the genus *Nasonia* (Brucker and Bordenstein 2012). There are only three species within this genus, and all three were used in this study: *N. vitripennis*, *N. giraulti*, and *N. longicornis* (Campbell et al. 1994). All wasps were collected from the wild and inbred in the laboratory to cure them of *Wolbachia* infection. This could lead to a lower estimate of bacterial diversity than would be expected in the wild. All three species were raised under identical conditions. Ten individuals were pooled together from each species and universal 16S primers were used during PCR to amplify the gene. The sequences were cloned and ~90 clones from each sample were Sanger sequenced. In the adult stage of this wasp, they found an average of 15.6 OTUs in each laboratory-reared wasp species. Furthermore, six OTUs were shared between each species on average (61.6%). This yields an average of 9.6 bacterial species per host species in the genus *Nasonia*.

Averaging the estimates of mean bacterial species per host species from these three insect host genera (13.4, 9, 9.6) yields a mean estimate of 10.7 bacterial species unique to each arthropod species. Unfortunately, comparisons of gut microbiomes between closely related species are lacking in other arthropod orders. However, the fact that bacterial richness in two of the most species-rich insect orders converge on similar values gives us some confidence that values should be similar in other major insect orders also. Furthermore, studies across insects (although not addressing host specificity) confirm that numerous bacterial species occur in each insect species, including in the most species-rich insect orders (e.g., Coleoptera, Diptera, Hymenoptera, Lepidoptera; Table S3). We note that bacterial species richness is considerably higher in the estimates of Colman et al. (2012) than those in Jones et al. (2013). For example, Jones et al. (2013) estimate seven, nine, eight, and seven bacterial species per host species for Coleoptera, Diptera, Hymenoptera, and Lepidoptera, whereas Colman et al. (2012) estimated 25, 28, 11, and 32 (respectively). Colman et al. (2012) generally examined more species per order, and used a greater diversity of molecular approaches to estimate bacterial richness.

Free-Living Bacteria

Several overviews have been published of free-living bacterial diversity in soil (Fierer and Jackson 2006; Lozupone and Knight 2007; Fierer et al. 2012), atmospheric (Bowers et al. 2009; Gandolfi et al. 2013), marine (Sogin et al. 2006; Pommier et al. 2007; Zinger et al. 2011), and deep-earth habitats (Fry et al. 2008; Griebler and Lueders 2009; Edwards et al. 2011). Details on bacterial diversity in each of these habitats are described below. Overall, a commonly cited figure for free-living bacteria is ~6 million species (Curtis et al. 2002).

bacteria in soils

Recent studies suggest that bacterial species richness at local scales in soil can be high. However, a reliable estimate of global diversity has been elusive, given that the amount of species turnover within and between local sites is unclear. Depending on the soil type, 10 gram soil

samples have contained between ~3000 and ~12,000 bacterial OTUs, based on 16S metagenomic sequences (Fierer et al. 2012). Furthermore, when bacterial endemism has been compared across soil samples, 74.4% of OTUs were found to be unique to a single soil sample (Fulthorpe et al. 2008). Nevertheless, microbial species diversity in the soil remains contentious (Dykhuizen 1998; Schloss and Handelsman 2004). Although a recent paper suggested there may be ~1 trillion microbial species, the environment occupied by these species is not known (parasitic versus free-living; Locey and Lennon 2016).

BACTERIA IN THE ATMOSPHERE

Bacteria in the atmosphere can be metabolically active and may play a role in weather events such as cloud formation (Gandolfi et al. 2013; Behzad et al. 2015). Bacteria also inhabit the upper troposphere, with sample locations at 10 km above the surface containing 99 to 299 bacterial OTUs per 6 m^3 of air (DeLeon-Rodriguez et al. 2013). However, these bacteria appear to match those from both soil and marine environments. Therefore, the atmosphere might have relatively limited numbers of unique bacterial species.

BACTERIA IN THE OCEAN

In a large analysis of marine bacterial diversity studies, a total of 9.6 million 16S sequences from 509 globally distributed marine samples recovered a total of ~120,000 OTUs at the 3% sequence identity cutoff (Zinger et al. 2011).

CAVEATS

Molecular studies of uncultivable bacteria have revealed high levels of diversity that were previously unknown. Next-generation sequencing technologies now allow the simultaneous amplification of millions of bacterial DNA sequences. Nevertheless, there are a few caveats. First, the use of "universal" 16S primers to amplify all bacterial sequences in a population for sequencing may insert bias in the estimates. These primers were used in the majority of the studies we reported here. One paper suggested that up to half of all bacterial diversity could be missed because these universal primers may fail to amplify many species (Hong et al. 2009). Second, bacterial species (OTUs) are typically recognized based on a criterion of clustering of 16S sequences at a 97% identity level. However, this criterion may be overly conservative, because it has been shown that different species of bacteria can harbor nearly identical 16S sequences and yet be differentiated phenotypically (Fox et al. 1992; Stackebrandt and Goebel 1994). In contrast, bacterial diversity may sometimes be overestimated due to PCR and sequencing error (Kunin et al. 2010). Rigorous controls need to be in place in order to assess the true diversity within samples.

REFERENCES

Amann R. I., Ludwig W., Schleifer K. H. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews* 59:143–169.

Behzad H., Gojobori T., Mineta K. 2015. Challenges and opportunities of airborne metagenomics. *Genome Biology and Evolution* 7:1216–1226.

Bowers R. M., Lauber C. L., Wiedinmyer C., Hamady M., Hallar A. G., Fall R., Knight R., Fierer N. 2009. Characterization of airborne microbial communities at a high-elevation site and their potential to act as atmospheric ice nuclei. *Applied and Environmental Microbiology* 75:5121–5130.

Brucker R. M., Bordenstein S. R. 2012. The roles of host evolutionary relationships (Genus: *Nasonia*) and development in structuring microbial communities. *Evolution* 66:349–362.

Campbell B. C., Steffen-Campbell J. D., Werren J. H. 1994. Phylogeny of the *Nasonia* species complex (Hymenoptera: Pteromalidae) inferred from an Internal Transcribed Spacer (ITS2) and 28S rDNA sequences. *Insect Molecular Biology* 2:225–237.

Chaisiri K., McGarry J. W., Morand S., Makepeace B. L. 2015. Symbiosis in an overlooked microcosm: a systematic review of the bacterial flora of mites. *Parasitology* 142:1152–1162.

Chan T.-F., Ji K.-M., Yim A. K.-Y., et al. 2015. The draft genome, transcriptome, and microbiome of *Dermatophagoides farinae* reveal a broad spectrum of dust mite allergens. *Journal of Allergy and Clinical Immunology* 135:539–548.

Chandler J. A., Morgan Lang J., Bhatnagar S., Eisen J. A., Kopp A. 2011. Bacterial communities of diverse *Drosophila* species: ecological context of a host-microbe model system. *PLOS Genetics* 7:e1002272.

Colman D. R., Toolson E. C., Takacs-Vesbach C. D. 2012. Do diet and taxonomy influence insect gut bacterial communities? *Molecular Ecology* 21:5124–5137.

Curtis T. P., Sloan W. T., Scannell J. W. 2002. Estimating prokaryotic diversity and its limits. *Proceedings of*

the National Academy of Sciences of the United States of America 99:10494–10499.

DeLeon-Rodriguez N., Lathem T. L., Rodriguez-R L. M., Barazesh J. M., Anderson B. E., Beyersdorf A. J., Ziemba L. D., Bergin M., Nenes A., Konstantinidis K. T. 2013. Microbiome of the upper troposphere: species composition and prevalence, effects of tropical storms, and atmospheric implications. *Proceedings of the National Academy of Sciences of the United States of America* 110:2575–2580.

Dykhuizen D. E. 1998. Santa Rosalia revisited: why are there so many species of bacteria? *Antonie van Leeuwenhoek* 73:25–33.

Edwards K. J., Wheat C. G., Sylvan J. B. 2011. Under the sea: microbial life in volcanic oceanic crust. *Nature Reviews Microbiology* 9:703–712.

Engel P., Moran N. A. 2013. The gut microbiota of insects—diversity in structure and function. *FEMS Microbiology Reviews* 37:699–735.

Fierer N., Jackson R. B. 2006. The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America* 103:626–631.

Fierer N., Leff J. W., Adams B. J., Nielsen U. N., Bates S. T., Lauber C. L., Owens S., Gilbert J. A., Wall D. H., Caporaso J. G. 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences of the United States of America* 109:21390–21395.

Fox G. E., Wisotzkey J. D., Jurtshuk P. Jr. 1992. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *International Journal of Systematic Bacteriology* 42:166–170.

Fry J. C., Parkes R. J., Cragg B. A., Weightman A. J., Webster G. 2008. Prokaryotic biodiversity and activity in the deep subseafloor biosphere. *FEMS Microbiology Ecology* 66:181–196.

Fulthorpe R. R., Roesch L. F. W., Riva A., Triplett E. W. 2008. Distantly sampled soils carry few species in common. *ISME Journal* 2:901–910.

Gandolfi I., Bertolini V., Ambrosini R., Bestetti G., Franzetti A. 2013. Unravelling the bacterial diversity in the atmosphere. *Applied Microbiology and Biotechnology* 97:4727–4736.

Griebler C., Lueders T. 2009. Microbial biodiversity in groundwater ecosystems. *Freshwater Biology* 54:649–677.

Hong S., Bunge J., Leslin C., Jeon S., Epstein S. S. 2009. Polymerase chain reaction primers miss half of rRNA microbial diversity. *ISME Journal* 3:1365–1373.

Hubert J., Kopecký J., Perotti M. A., Nesvorná M., Braig H. R., Ságová-Marečková M., Macovei L., Zurek L. 2012. Detection and identification of species-specific bacteria associated with synanthropic mites. *Microbial Ecology* 63:919–928.

Jones R. T., Sanchez L. G., Fierer N. 2013. A cross-taxon analysis of insect-associated bacterial diversity. *PLOS ONE* 8:e61218.

Kunin V., Engelbrektson A., Ochman H., Hugenholtz P. 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology* 12:118–123.

Ladygina N., Johansson T., Canbäck B., Tunlid A., Hedlund K. 2009. Diversity of bacteria associated with grassland soil nematodes of different feeding groups. *FEMS Microbiology Ecology* 69:53–61.

Ley R. E., Hamady M., Lozupone C., Turnbaugh P. J., Ramey R. R., Bircher J. S., Schlegel M. L., Tucker T. A., Schrenzel M. D., Knight R., Gordon J. I. 2008. Evolution of mammals and their gut microbes. *Science* 320:1647–1651.

Locey K. J., Lennon J. T. 2016. Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences of the United States of America* 113:5970–5975.

Lozupone C. A., Knight R. 2007. Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences of the United States of America* 104:11436–11440.

Muegge B. D., Kuczynski J., Knights D., Clemente J. C., González A., Fontana L., Henrissat B., Knight R., Gordon J. I. 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* 332:970–974.

Nemergut D. R., Costello E. K., Hamady M., Lozupone C., Jiang L., Schmidt S. K., Fierer N., Townsend A. R., Cleveland C. C., Stanish L., Knight R. 2011. Global patterns in the biogeography of bacterial taxa. *Environmental Microbiology* 13:135–144.

Pace N. R. 1997. A molecular view of microbial diversity and the biosphere. *Science* 276:734–740.

Pommier T., Canbäck B., Riemann L., Boström K. H., Simu K., Lundberg P., Tunlid A., Hagström A. 2007. Global patterns of diversity and community structure in marine bacterioplankton. *Molecular Ecology* 16:867–880.

Sanders J. G., Powell S., Kronauer D. J. C., Vasconcelos H. L., Frederickson M. E., Pierce N. E. 2014. Stability and phylogenetic correlation in gut microbiota: lessons from ants and apes. *Molecular Ecology* 23:1268–1283.

Schloss P. D., Handelsman J. 2004. Status of the microbial census. *Microbiology and Molecular Biology Reviews* 68:686–691.

Schmitt S., Tsai P., Bell J., Fromont J., Ilan M., Lindquist N., Perez T., Rodrigo A., Schupp P. J., Vacelet J., Webster N., Hentschel U., Taylor M. W. 2012. Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *ISME Journal* 6:564–576.

Sogin M. L., Morrison H. G., Huber J. A., Welch D. M., Huse S. M., Neal P. R., Arrieta J. M., Herndl G. J.

2006. Microbial diversity in the deep sea and the underexplored "rare biosphere." *Proceedings of the National Academy of Sciences of the United States of America* 103:12115–12120.

Stackebrandt E., Goebel B. M. 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology* 44:846–849.

Sullam K. E., Essinger S. D., Lozupone C. A., O'Connor M. P., Rosen G. L., Knight R., Kilham S. S., Russell J. A. 2012. Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. *Molecular Ecology* 21:3363–3378.

Torsvik V., Goksøyr J., Daae F. L. 1990. High diversity in DNA of soil bacteria. *Applied and Environmental Microbiology* 56:782–787.

Waite D. W., Taylor M. W. 2014. Characterizing the avian gut microbiota: membership, driving influences, and potential function. *Frontiers in Microbiology* 5:223.

Whitman W. B., Coleman D. C., Wiebe W. J. 1998. Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences of the United States of America* 95:6578–6583.

Yun J.-H., Roh S. W., Whon T. W., Jung M.-J., Kim M.-S., Park D.-S., Yoon C., Nam Y.-D., Kim Y.-J., Choi J.-H., Kim J.-Y., Shin N.-R., Kim S.-H., Lee W.-J., Bae J.-W. 2014. Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. *Applied and Environmental Microbiology* 80:5254–5264.

Zindel R., Ofek M., Minz D., Palevsky E., Zchori-Fein E., Aebi A. 2013. The role of the bacterial community in the nutritional ecology of the bulb mite *Rhizoglyphus robini* (Acari: Astigmata: Acaridae). *FASEB Journal* 27:1488–1497.

Zinger L., Amaral-Zettler L. A., Fuhrman J. A., Horner-Devine M. C., Huse S. M., Welch D. B. M., Martiny J. B. H., Sogin M., Boetius A., Ramette A. 2011. Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *PLOS ONE* 6:e24570.

TABLE S1

Summary of estimated numbers of cryptic species per morphology-based described species in arthropods, averaged across studies

	Included studies	Mean cryptic species per described species	Standard deviation
Insects	7	5.95	3.62
Non-insect arthropods	9	5.84	5.42
Total	16	5.89	4.57

Full data are in Table S2. See Appendix S1 for discussion.

TABLE S2

Number of cryptic species delimited per named (described) species based on Bayesian species delimitation (BPP), including mean and standard deviation per study

Focal taxon	Cryptic species delimited per described species	Study mean and standard deviation (described species considered)	Reference
	Insects		
Asecodes parasitoid wasps	4 in 1	4 (1)	Hamback et al. (2013)
Bemisia whiteflies	9 in 1	9 (1)	Hsieh et al. (2014)
Pyrgonota leafhoppers	9 in 1	9 (1)	Su et al. (2014)
Scirtothrips thrips	9 in 1	9 (1)	Dickey et al. (2015)
Sphenarium grasshoppers	2 in 1, 2 in 1; cryptic lineages not found in 2 species	1.5 ± 0.58 (4)	Pedraza-Lara et al. (2015)
Polyura butterflies	3 in 1, 2 in 1, 2 in 1, 2 in 1; cryptic lineages not found in 28 species	1.2 ± 0.448 (32)	Toussaint et al. (2015)
Cicadetta cicadas	8 in 1	8 (1)	Wade et al. (2015)
	Non-insect arthropods		
Aliatypus trapdoor spiders	6 in 1	6 (1)	Satler et al. (2013)
Nesticella cave spiders	2 in 1, 3 in 1, 2 in 1; cryptic lineages not found in 5 species	1.5 ± 0.76 (8)	Zhang and Li (2013)
Titanidiops trapdoor spiders	2 in 1	2 (1)	Opatova and Arnedo (2014)
Tomocerus springtails	19 in 1, 7 in 1	13 ± 8.49 (2)	Zhang et al. (2014)
Telema cave spiders	16 in 1	16 (1)	Zhang and Li (2014)
Cicurina cave spiders	3 named spiders delimited as 1 species; 1 additional species with no cryptic lineages found	0.50 ± 0.34 (4)	Hedin (2015)
Microhexura moss spiders	6 in 1	6 (1)	Hedin et al. (2015)
Wangiannachiltonia amphipods	6 in 1	6 (1)	Murphy et al. (2015)
Megabunus harvestmen	3 in 1, 2 in 1; cryptic lineages not found in 3 species	1.6 ± 0.89 (5)	Wachter et al. (2015)

Note that we did not include new species not previously identified to named species. In the study by Wade et al. (2015), there were 13 named species included in the BPP analysis, but these were described based on song differences. Given our focus on morphologically defined species, we considered these all to be one species (*Cicadetta montana*), in accordance with previous, morphology-based taxonomy. In the study by Opatova and Arnedo (2014), we used the most conservative estimated number of cryptic species across different analyses. See Appendix S1 for discussion and references.

TABLE S3

Species richness of gut-associated bacteria across animal hosts

Phylum	Class	Order	Mean bacterial OTUs/host	Host species	Sequencing method	Source of host	Reference
Arthropoda	Arachnida	Sarcoptiformes (mites)	10	4	Sanger	Lab cultured	Hubert et al. (2012)
Arthropoda	Arachnida	Sarcoptiformes (mites)	100	1	WGS	Lab cultured	Chan et al. (2015)
Arthropoda	Arachnida	Sarcoptiformes (mites)	557	1	454	Lab cultured	Zindel et al. (2013)
Arthropoda	Insecta	Coleoptera	7	6	454	Wild	Jones et al. (2013)
Arthropoda	Insecta	Diptera	9	13	454	Wild	Jones et al. (2013)
Arthropoda	Insecta	Hemiptera	7	8	454	Wild	Jones et al. (2013)
Arthropoda	Insecta	Hymenoptera	8	7	454	Wild	Jones et al. (2013)
Arthropoda	Insecta	Isoptera	6	1	454	Wild	Jones et al. (2013)
Arthropoda	Insecta	Lepidoptera	7	1	454	Wild	Jones et al. (2013)
Arthropoda	Insecta	Neuroptera	2	1	454	Wild	Jones et al. (2013)
Arthropoda	Insecta	Siphonaptera	6	2	454	Wild	Jones et al. (2013)
Arthropoda	Insecta	Hymenoptera	11	20	Sanger/454	Wild	Colman et al. (2012)
Arthropoda	Insecta	Coleoptera	25	19	Sanger/454	Wild	Colman et al. (2012)
Arthropoda	Insecta	Diptera	28	4	Sanger/454	Wild	Colman et al. (2012)
Arthropoda	Insecta	Isoptera	95	12	Sanger/454	Wild	Colman et al. (2012)
Arthropoda	Insecta	Lepidoptera	32	3	Sanger/454	Wild	Colman et al. (2012)
Arthropoda	Insecta	Neuroptera	8	1	Sanger/454	Wild	Colman et al. (2012)
Arthropoda	Insecta	Heteroptera	3	1	Sanger/454	Wild	Colman et al. (2012)
Chordata	Mammalia	Artiodactyla	102 (54.8)	11	Sanger	Wild/Captive	Colman et al. (2012)
Chordata	Mammalia	Carnivora	35 (47.3)	9	Sanger	Captive	Colman et al. (2012)
Chordata	Mammalia	Chiroptera	63 (42.1)	2	Sanger	Captive	Colman et al. (2012)
Chordata	Mammalia	Hyracoidea	43 (53.1)	1	Sanger	Captive	Colman et al. (2012)
Chordata	Mammalia	Insectivora	54 (55.6)	1	Sanger	Captive	Colman et al. (2012)
Chordata	Mammalia	Lagomorpha	54 (87.0)	1	Sanger	Captive	Colman et al. (2012)
Chordata	Mammalia	Perissodactyla	136 (64.1)	6	Sanger	Wild/Captive	Colman et al. (2012)
Chordata	Mammalia	Primates	98 (59.0)	16	Sanger	Wild/Captive	Colman et al. (2012)
Chordata	Mammalia	Proboscidea	87 (56.9)	2	Sanger	Wild/Captive	Colman et al. (2012)
Chordata	Mammalia	Rodentia	108 (80.8)	3	Sanger	Captive	Colman et al. (2012)
Chordata	Mammalia	Xenarthra	94 (60.6)	1	Sanger	Captive	Colman et al. (2012)
Chordata	Mammalia	Diprotodontia	111 (66.4)	1	Sanger	Captive	Colman et al. (2012)

Chordata	Mammalia	Monotremata	68 (82.4)	1	Sanger	Captive	Colman et al. (2012)
Chordata	Mammalia		531	39	454	Wild/Captive	Muegge et al. (2011)
Chordata	Actinopterygii (Fish)	Perciformes	34	9	Sanger	Wild/Captive	Sullam et al. (2012)
Chordata	Actinopterygii (Fish)	Cypriniformes	85	1	Sanger	Wild/Captive	Sullam et al. (2012)
Chordata	Actinopterygii (Fish)	Gasterosteiformes	39	1	Sanger	Captive	Sullam et al. (2012)
Chordata	Actinopterygii (Fish)	Salmoniformes	16	4	Sanger	Wild/Captive	Sullam et al. (2012)
Chordata	Actinopterygii (Fish)	Siluriformes	52	1	Sanger	Wild	Sullam et al. (2012)
Chordata	Actinopterygii (Fish)	Cyprinodontiformes	70	1	Sanger	Wild	Sullam et al. (2012)
Chordata	Actinopterygii (Fish)	Pleuronectiformes	18	1	Sanger	Captive	Sullam et al. (2012)
Chordata	Actinopterygii (Fish)	Tetraodontiformes	21	1	Sanger	Wild	Sullam et al. (2012)
Chordata	Aves	Sphenisciformes	44	1	Sanger	N/A	Waite and Taylor (2014)
Chordata	Aves	Galliformes	126	3	Sanger	N/A	Waite and Taylor (2014)
Chordata	Aves	Gruiformes	7	1	Sanger	N/A	Waite and Taylor (2014)
Chordata	Aves	Opisthocomiformes	322	1	Sanger	N/A	Waite and Taylor (2014)
Chordata	Aves	Psittaciformes	12	1	Sanger	N/A	Waite and Taylor (2014)
Chordata	Aves	Charadriiformes	85	1	Sanger	N/A	Waite and Taylor (2014)
Chordata	Aves	Struthioniformes	98	1	Sanger	N/A	Waite and Taylor (2014)
Nematoda	Secernentea	Rhabditida	35	9	Sanger	Wild	Ladygina et al. (2009)
Porifera	Demospongiae	Astrophorida	261	5	454	Wild	Schmitt et al. (2012)
Porifera	Demospongiae	Hadromerida	45	1	454	Wild	Schmitt et al. (2012)
Porifera	Demospongiae	Verongida	138	4	454	Wild	Schmitt et al. (2012)
Porifera	Demospongiae	Dictyoceratida	146	8	454	Wild	Schmitt et al. (2012)
Porifera	Demospongiae	Haplosclerida	199	3	454	Wild	Schmitt et al. (2012)
Porifera	Demospongiae	Halichondrida	32	2	454	Wild	Schmitt et al. (2012)
Porifera	Demospongiae	Chondrosida	42	1	454	Wild	Schmitt et al. (2012)
Porifera	Homoscleromorpha	Homosclerophorida	77	2	454	Wild	Schmitt et al. (2012)
Porifera	Demospongiae	Poecilosclerida	111	1	454	Wild	Schmitt et al. (2012)
Porifera	Demospongiae	Lithistida	131	1	454	Wild	Schmitt et al. (2012)

All studies used the 16S gene (either in part or full), and an identity threshold of 96 or 97% for considering sequence clusters to be unique species. The number of operational taxonomic units (OTUs) is used as a proxy for species. The mean % of unique species is given in parentheses with the mean bacterial species per host species, and refers to OTUs that are unique to that order (mostly in mammals). The number of host species refers to the number of sampled species within that host order that were used in calculating the mean number of bacterial OTUs per host. The sequencing methods included Sanger sequencing of clones, 454 pyrosequencing, or WGS (whole-genome shotgun sequencing). See Appendix S5 for references.