Estimating Global Biodiversity: The Role of Cryptic Insect Species

Xin Li[1] and John J. Wiens[2,*]

[1]Department of Entomology, China Agricultural University, Beijing 100193, China
[2]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721-0088, USA
[*]Correspondence to be sent to: Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ. 85721-0088, USA; E-mail:
wiensj@email.arizona.edu

Abstract.—How many species are there on Earth and to what groups do these species belong? These fundamental questions span systematics, ecology, and evolutionary biology. Yet, recent estimates of overall global biodiversity have ranged wildly, from the low millions to the trillions. Insects are a pivotal group for these estimates. Insects make up roughly half of currently described extant species (across all groups), with ~1 million described species. Insect diversity is also crucial because many other taxa have species that may be unique to each insect host species, including bacteria, apicomplexan protists, microsporidian fungi, nematodes, and mites. Several projections of total insect diversity (described and undescribed) have converged on ~6 million species. However, these projections have not incorporated the morphologically cryptic species revealed by molecular data. Here, we estimate the extent of cryptic insect diversity. We perform a systematic review of studies that used explicit species-delimitation methods with multilocus data. We estimate that each morphology-based insect species contains (on average) 3.1 cryptic species. We then use these estimates to project the overall number of species on Earth and their distribution among major groups. Our estimates suggest that overall global biodiversity may range from 563 million to 2.2 billion species. [Biodiversity; cryptic species; insects; species delimitation; species richness.]

The estimation of global biodiversity represents a colossal failure of science. As astronomers and astrobiologists search for evidence of life on other worlds, one might assume that we knew roughly how many species are on our own planet, and to what groups those species belong (the Pie of Life; Larsen et al. 2017). The range of recent estimates of global biodiversity shows that we clearly do not know. For example, some authors have estimated that there are only ~2 million species on Earth (e.g., Costello et al. 2012), effectively the same as the current number of described species (2.01 million; Bánki et al. 2021). Others have estimated that there are a trillion species or more (Locey and Lennon 2016). A very well-cited study has suggested that there are ~11 million extant species on Earth, with bacteria making up a negligible portion of those species (Mora et al. 2011). Other studies have estimated vast numbers of microbial species (Locey and Lennon 2016), implicitly making macroscopic species richness trivial in comparison.

This uncertainty is far from strictly academic. For example, a recent report suggested that a million species are now threatened with extinction from human activities (Tollefson 2019). But that estimate depended on a particular projection of overall global biodiversity, one much larger than the number of described species but much smaller than other estimates.

Insects are a pivotal group for estimating global biodiversity. First, insects currently make up roughly half of all extant, described species on Earth, when all species are considered across all groups (Bánki et al. 2021). Given this, it is reasonable to assume that they will also be influential when undescribed species are considered. Second, there is evidence that insects may be host to numerous other types of organisms (as parasites, mutualists, or commensals). For example, a recent review (Larsen et al. 2017) suggested that each insect species may host (on average) a unique species of mite, nematode, microsporidian fungus, and apicomplexan protist, and ~11 bacterial species (or ~8; Wiens 2021). This host-associated diversity may dwarf free-living diversity in the most species-rich groups (i.e., animals, fungi, protists, bacteria). Thus, the overall number of species on Earth (across all groups) may hinge on the number of insect species.

There is both good news and bad news about projections of total insect diversity (with "total" including both described and undescribed species). The good news is that projections of insect richness have been surprisingly consistent across studies for decades. Specifically, several analyses and reviews have concluded that there are roughly 6 million insect species, including Gaston (1991), Hammond (1995), Groombridge and Jenkins (2002), Novotny et al. (2002), Grimaldi and Engel (2005), Raven and Yeates (2007), Chapman (2009), Mora et al. (2011), Basset et al. (2012), and Stork et al. (2015). These studies used very different approaches to estimate these numbers (e.g., expert opinion of insect systematists, insect–plant associations, ratios of richness among taxonomic ranks, body size and year of description, ratios of known to unknown species). Nevertheless, they arrived at broadly similar conclusions, especially given the huge range of overall estimates of biodiversity when all groups are considered (from millions to trillions; see above).

The bad news is that these projections of global insect diversity did not include morphologically cryptic

species revealed by molecular data. Cryptic species have been widely documented in numerous groups of organisms, and may have important consequences for estimating global biodiversity (e.g., Bickford et al. 2007; Pfenninger and Schwenk 2007; Adams et al. 2014; Singhal et al. 2018; Struck et al. 2018). Unfortunately, these previous projections of insect diversity assumed (either implicitly or explicitly) that morphological data alone were sufficient to reveal the overall number of insect species.

Larsen et al. (2017) tried to address this issue by explicitly incorporating cryptic insect species in their projections of global biodiversity. They conducted a systematic search for studies (published from 2013 to 2015) that used multilocus data to delimit species in arthropods. They specifically focused on studies that used the Bayesian phylogenetics and phylogeography (BPP) method (Yang and Rannala 2010) for species delimitation. The BPP method is widely used and often considered to be relatively accurate (e.g., Camargo et al. 2012; Rannala 2015), despite some controversy (Sukumaran and Knowles 2017; Leaché et al. 2019). For each of 16 studies found in their literature review, they counted the number of unnamed species supported by BPP relative to the number of described (morphology-based) species included in the study. Based on averages across each study, they inferred that each morphology-based species contained approximately 6 cryptic species (on average). Thus, rather than there being ~6 million insect species, these estimates would suggest that ~36 million insect species might be more likely.

There are some potential problems with the estimates of cryptic insect richness in Larsen et al. (2017). First, "cryptic species" was included as a keyword in the literature searches. This might have biased the results to find more cryptic species and thereby overestimate their frequency. Larsen et al. (2017) discussed this potential source of bias, and noted that many studies found were focused on phylogeography and not cryptic species at all. Nevertheless, this is a crucial issue to address. Second, the number of insect studies included was limited (only 8). Larsen et al. (2017) also included 8 studies of other arthropods (mostly spiders), which yielded similar results. A secondary problem arising from the limited sampling is that a simple mean across studies may not be fully accurate for projecting overall insect diversity. For example, given a limited number of studies, the estimates of cryptic species might (by chance) reflect values from groups that are not broadly representative of insects overall. Specifically, in the survey of Larsen et al. (2017), nearly half of the insect studies were of hemipterans (the fifth largest insect order in terms of species richness), but none were of coleopterans (the largest, about 4 times larger than Hemiptera; Bánki et al. 2021). A better approach would be to estimate the average number of cryptic species (per morphology-based species) for each insect order, and weight the overall estimate of cryptic insect diversity by the relative richness of each order.

Given these issues, a re-estimation of morphologically cryptic insect diversity is urgently needed. This is crucial not only for understanding overall insect richness, but also for understanding global biodiversity overall. Here, we provide such a re-estimate. We start from a less restrictive search for case studies. We generate a much larger sample size of studies, and then weight the overall estimates of cryptic diversity within each insect order by the richness of those orders. We address numerous potential sources of bias in these estimates (which were not addressed previously), including sampling of individuals within species, different species-delimitation methods, sampling of genera across latitudes, and sampling of species for inclusion in species-delimitation studies. We then use these estimates of cryptic insect species to project overall global insect richness, and the richness of other groups as well (all animals, plants, fungi, bacteria, and protists), given the numerous taxa that are associated with insect hosts. Our results support the idea that there are likely to be multiple cryptic species for each morphology-based species (roughly 3 on average), but not as many as estimated by Larsen et al. (2017). These new estimates also allow us to substantially narrow the broad range of estimates for global biodiversity.

Materials and Methods

Estimating Numbers of Cryptic Insect Species

Our overall goal was to estimate the number of species inferred from molecular data relative to the number of morphology-based species across insects, and then use this number to project overall global diversity of insects and other groups. To do this, we needed to find case studies that performed molecular-based species-delimitation analyses, with case studies spanning many different insect genera and orders. For each genus that was the subject of a species-delimitation study, we sought the ratio of the number of inferred molecular-based species to the number of morphology-based species that were initially included. A ratio > 1 indicates that there are 1 or more morphologically cryptic species in the genus. On the other hand, a ratio < 1 indicates that the molecular species-delimitation analyses found that 1 or more morphology-based species were not actually distinct. Cryptic species were defined as 2 or more distinct species that are erroneously classified (and hidden) under one species name based on morphological data. These species could differ in other phenotypic characteristics (e.g., ecology, mating signals), and there might be various reasons why these species were overlooked morphologically (Struck et al. 2018). Note that inferences of species limits based on molecular data generally supported those based on morphology, but with additional species often recognized within one or more morphology-based species (i.e., the cryptic, molecular-based species). Therefore,

we emphasized the ratio of molecular-based to morphology-based species.

We searched Google Scholar with the keywords "phylogeography, insect, nuclear, species delimitation" on September 12, 2020. The phrase "cryptic species" was not used, to avoid biasing the results to favor finding more cryptic species. We assumed that phylogeographic studies should be neutral about the presence of cryptic species (yet should also have the potential to reveal them). Therefore, the keyword "phylogeography" was used. We included the keyword "nuclear" to target studies that included nuclear genes and not only mitochondrial genes. Many articles found only focused on phylogeographic questions. The keyword "species delimitation" was used to help target studies that performed species-delimitation analyses (as opposed to merely analyzing phylogeography, estimating phylogeny, or other topics).

A total of approximately 3,650 results were found in this initial search. These results were then sorted by relevance. When we reached a set of 60 articles in a row in which the titles and abstracts were not related to our keywords, we assumed that subsequent articles would also not be relevant (given the sorting by relevance). This left 269 papers that were potentially relevant. We then further screened this set of 269 papers. Ninety-one of these 269 articles were excluded because they did not focus on species delimitation in insects. We also excluded analyses of non-native species that were sampled only in their introduced ranges, since these analyses may underestimate cryptic species diversity relative to analyses of the species' native range. Articles were also excluded if they did not include molecular data or focused only on phylogeny (with no species delimitation).

Our focus was on estimating the number of morphology-based species relative to the number of molecular-based species to calculate the ratio of cryptic species to morphology-based species. Papers were therefore excluded if the number of morphology-based species sampled in the genus was not clear. Similarly, we excluded studies in which the number of species estimated by the species-delimitation methods was not explicitly given. A list of the methods used for species delimitation (and their abbreviations) is given in Supplementary Table S1 (all supplementary tables and datasets are available on Dryad: http://dx.doi.org/10.5061/dryad.vt4b8gtt7).

Our primary data set included only studies that delimited species using nuclear and mitochondrial data (most studies) or nuclear data alone (Supplementary Table S2). We also created a separate data set (Supplementary Table S3) for studies found in our search that only delimited species using mitochondrial data. A larger pool of mitochondrial-only studies could presumably have been found if our searches had not included the keyword "nuclear." Our emphasis here on multilocus studies (nuclear+mitochondrial) over single-locus studies (mitochondrial only) should

be uncontroversial. Indeed, recent studies have highlighted the potential for overestimation of cryptic species diversity using mitochondrial data alone (e.g., Chan et al. 2022).

Some studies were ambiguous as to what category they belonged to. Jin et al. (2018) used mitochondrial data to delimit species and also re-analyzed a genus using both mitochondrial and nuclear genes. Because the samples and results were independent between the 2 steps, we counted them separately in the 2 data sets (nuclear vs. mitochondrial only). In Zhu et al. (2017), only analyses with the mitochondrial genes were counted because the nuclear gene was from a bacterial aphid endosymbiont.

Whenever possible, we followed the authors' conclusions about the overall number of cryptic species they inferred, if this was given. However, most studies did not explicitly state this number, and most studies used more than one method to delimit species (see Supplementary Tables S2 and S3). In these cases, we used the number of inferred species that was supported most frequently by different methods (i.e., the majority or plurality). If there was no majority or plurality (e.g., only 2 methods were used and they inferred different numbers), then the mean number of inferred species across all analyses was used instead. All separate species-delimitation analyses performed by the authors were counted here. The different analyses depended on the study (e.g., different genes and their combinations, different delimitation thresholds, different sexes). To evaluate whether the overall results were sensitive to the use of different methods within each study, we performed a set of analyses in which we used the minimum number of species estimated in each study, and another set in which we used the maximum number.

Some papers used genetic methods to delimit species that were not explicitly designed for species delimitation (Supplementary Table S1). These included Networks (Bandelt et al. 1999) and Structure (Pritchard et al. 2000; Falush et al. 2003). We also incorporated the results of these analyses, when the authors were explicit about how many species they had inferred from these methods.

We also obtained the number of individuals sampled in each study for molecular data (Supplementary Tables S4 and S5). This was important in order to evaluate whether some studies found few cryptic species simply because they sampled few individuals per morphology-based species (see below). Outgroup individuals were excluded from these counts because outgroup species were also excluded in our estimates of cryptic species. We followed the authors' statements regarding how many individuals were used (e.g., if different numbers were given in Methods, Results, and/or Supplementary Materials). Small discrepancies in these numbers should have little impact on our conclusions.

Our initial analyses revealed that cryptic species were only rarely found in those studies in which relatively few individuals were sampled for molecular data

for each morphology-based species (Supplementary Table S6). In a regression analysis of the data in Supplementary Table S6 (Fig. 1), we found a strong, significant relationship among genera between the mean number of individuals sampled per morphology-based species in each genus and the mean number of cryptic species inferred per morphology-based species in that genus ($r^2 = 0.316$, $P < 0.0001$, $n = 45$). This relationship was strong because in most genera in which relatively few individuals were sampled per morphology-based species, very few cryptic species were generally found. By contrast, studies with better sampling of individuals could either infer few cryptic species per morphology-based species or many (this variability helps explain why this relationship is not perfect; Fig. 1). Therefore, to avoid artifacts associated with limited sampling, we excluded genera in which 10 or fewer individuals were sampled per morphology-based species (on average). The value of 10 is arbitrary. However, we preferred to err on the side of excluding studies, rather than including those with potentially problematic sample sizes. These were our primary analyses. We also performed supplementary analyses in which we evaluated the effects of using other cutoffs, specifically 5 and 20 (i.e., excluding studies in which 5 individuals or fewer were sampled, or 20 individuals or fewer, on average). However, a cutoff of 5 is clearly problematic, since a single, well-sampled morphology-based species can have >5 cryptic species (see Results).

   We also tested whether there was a strong latitudinal effect on the ratio of cryptic species, given that such an effect has been suggested (e.g., Freeman and Pennell 2021). We first estimated the approximate latitudinal position of each species in the 25 genera that were the focus of the primary analysis (Table 1). For most genera (22 of 25), we used locality information from the

original studies. However, the authors did not necessarily provide the latitude of each locality. In these cases, we searched for the coordinates of named localities in GoogleMaps and Baidu Map. For samples with unclear locations that were too broad (e.g., a country name only), we searched for additional localities in GBIF (www.gbif.org) using the R package *rgbif* (Chamberlain and Boettiger 2017). We also searched for localities for each species that lacked georeferenced localities in the original study. We retrieved 93,647 occurrences records (GBIF.org 2022). We then cleaned these records using the following steps. First, we filtered the dataset by matching the species associated with the occurrence records with species name of the desired species. Second, we removed obviously invalid geographic coordinates (i.e., zero for both longitude and latitude, ~500 km away from the closest mainland). We appended those species with localities from GBIF to those from the original reference. In total, we obtained coordinate data for 93,349 samples, including 2013 from the original references (for 22 of 25 genera) and 91,336 localities from GBIF. The mean for each species is given in Dataset S1, and the full set of localities is given in Dataset S2 (available on Dryad).

   To explore the relationship between the latitudinal position of each genus and its ratio of molecular to morphology-based species, we first calculated the mean latitude among the localities within each species. We then estimated the mean latitude for each genus based on the mean latitude among the species within that genus. We converted the mean latitude of each genus to its absolute value (since we are interested in whether genera occur nearer or farther from the equator, not whether they occur in the northern or southern hemispheres, and negative values would interfere with a potential linear relationship). Finally, we performed ordinary least squares regression between the ratio of molecular:morphology-based species in each genus and the mean latitudinal position of the genus. We utilized the R package *ggplot2* to visualize our results (Wickham 2016).

   We note several caveats about this analysis. First, our values for species reflect where most sampled localities are, and our values for genera reflect both this and where most sampled species occur. Although the exact values could change using other methods to estimate these values, the methodology used here should reflect whether a genus occurs predominantly at higher or lower latitudes. Second, we did not perform a phylogenetic correction, because we lacked a genus-level phylogeny for these taxa. A phylogenetic correction would likely make significant values non-significant (which is not an issue here, see Results). Third, we acknowledge that differences among clades might make a latitudinal pattern across orders more difficult to find. In some ways, our test addresses whether there is a significant latitudinal pattern that is stronger than other factors that might influence the ratio of molecular to morphology-based species.
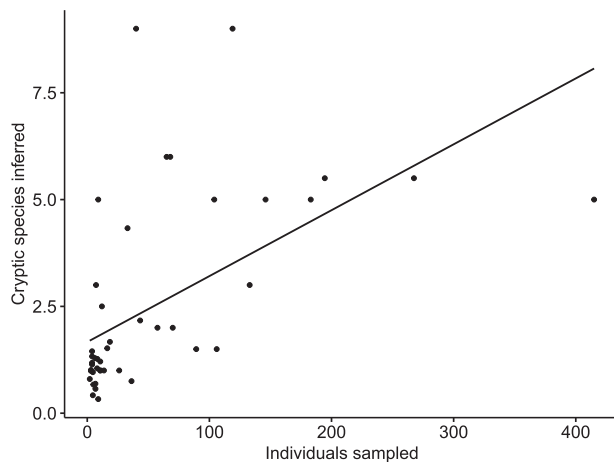

FIGURE 1.   Regression analysis between the number of individuals sampled (for molecular data) per morphology-based species and the estimated number of cryptic species inferred per morphology-based species. Each data point is a genus (not a study; some studies have multiple genera). Data are in Supplementary Table S6.

TABLE 1. Number of morphology-based and cryptic (molecular-based) species in each study

Order and family	Genus	Sampling	Ratio	Total species	Cryptic	Morph.	Reference
Coleoptera							
Cetoniidae	*Cetonia*	57.50	2	8	4	4	Ahrens et al. (2013)
Curculionidae	*Pachyrhynchus*	106	1.50	3	1	2	Chen et al. (2017)
Hydrophilidae	*Hydrobius*	194.50	5.50	11	9	2	Fossen et al. (2016)
Lampyridae	*Pteroptyx*	26.17	1	6	1*	6	Jusoh et al. (2020)
Diptera							
Chironomidae	*Tanytarsus*	12	2.50	15	8*	6	Lin et al. (2018)
Syrphidae	*Merodon*	13.67	1	3	0	3	Popović et al. (2015)
Ephemeroptera							
Baetidae	*Cloeon*	43.33	2.17	6.50	3.5	3	Rutschmann et al. (2017)
Hemiptera							
Aleyrodidae	*Bemisia*	40	9	9	8	1	Hsieh et al. (2014)
		9	5	5	4	1	de Moya et al. (2019)
Coccidae	*Parasaissetia*	65	6	6	5	1	Lin et al. (2017)
Diaspididae	*Chionaspis*	183	5	10	8	2	Gwiazdowski et al. (2011)
Eriococcidae	*Apiomorpha*	104	5	5	4	1	Cook and Rowell (2007)
Hymenoptera							
Agaonidae	*Pediobius*	146	5	5	4	1	Hernández-López et al. (2012)
Agaonidae	*Pleistodontes*	415	5	5	4	1	Darwell et al. (2014)
Andrenidae	*Andrena*	10.65	1.21	14.50	1*	12	Gueuning et al. (2020)
Halictidae	*Lasioglossum*	10.65	1	3	0	3	Gueuning et al. (2020)
Apidae	*Nomada*	10.65	1	2	0	2	Gueuning et al. (2020)
	Bombus	70	2	2	1	1	Martinet et al. (2018)
Braconidae	*Stenocorse*	119	9	9	8	1	Delgado-Machuca et al. (2020)
Formicidae	*Cataglyphis*	89.25	1.50	6	2	4	Eyer and Hefetz (2018)
		36.25	0.75	6	0	8	Eyer et al. (2017)
	Ectatomma	133	3	3	2	1	Aguilar-Velasco et al. (2016)
	Solenopsis	68	6	6	5	1	Ross et al. (2010)
Mecoptera							
Panorpidae	*Dicerapanorpa*	16.38	1.52	19.75	3*	13	Hu et al. (2019)
Orthoptera							
Lentulidae	*Betiscoides*	33	4.33	13	10	3	Matenaar et al. (2018)
Pyrgomorphidae	*Sphenarium*	18.56	1.67	15	6	9	Pedraza-Lara et al. (2015)
Thysanoptera							
Thripidae	*Scirtothrips*	267.50	5.50	11	9	2	Dickey et al. (2015)

Notes: These studies delimited species using nuclear data only or using both nuclear and mitochondrial data. Studies with mean molecular sampling of <10 individuals per morphology-based species were excluded. The data and species-delimitation method used in each study are summarized in Supplementary Table S2. Studies are listed alphabetically by order and then by family and genus within each order. "Sampling" was the total number of individuals sampled for molecular data, divided by the number of morphology-based species in the study (Supplementary Table S4). "Ratio" is the ratio of species inferred by molecular data to morphology-based species for each genus (molecular-based/morphology-based). "Total species" is the overall number of species estimated by molecular data. "Cryptic" is the overall estimated number of cryptic species. "Morph." is the number of species based on morphological data alone. The total number of species can be different from that obtained by simply adding the number of cryptic species and morphology-based species. Within a given study, the total number of species inferred most frequently by different species-delimitation methods was used (see Methods and Supplementary Table S3), but "*" indicates that the overall inferred number of cryptic species was stated by authors, and this number was used instead.

Projecting Global Biodiversity

We used the estimated numbers of molecular-based species per morphology-based species to project the total number of insect species on Earth and overall global biodiversity across groups. We did this in 3 steps, which we outline here and then describe in detail below. We first projected the total number of insect species by extrapolating the inferred ratios of molecular-based to morphology-based species to all described insect species. We then incorporated projections of undescribed insect species richness (which were based on morphology-based species) to estimate the total number of insect species, including described and undescribed species and morphology- and molecular-based species. Finally, we made projections of overall global biodiversity by incorporating groups with many insect-associated species, such as bacteria,

apicomplexan protists, microsporidian fungi, nematodes, and mites (following Larsen et al. 2017; Wiens 2021).

First, we estimated the number of molecular-based species per morphology-based species among all described insect species. To do this, we started with an estimate of the number of molecular-based species per morphology-based species for each sampled genus (Table 1). We then averaged values among genera for each insect order represented (Table 2). Again, we excluded genera in which <10 individuals were sampled for molecular data per morphology-based species. We then multiplied the mean ratio of molecular-based to morphology-based species among sampled genera in each order by the overall number of described species in that order (Table 2). We used the Catalogue of Life (CoL; Bánki et al. 2021) to obtain the current number of described species in each sampled

TABLE 2.   The mean ratio of molecular to morphology-based species, averaged among genera within each order

Order	Number of sampled genera	Mean ratio of molecular:morphology-based species among genera	Described species in order	Projected species richness
Coleoptera	4	2.50	277,586	693,965
Diptera	2	1.75	165,144	289,002
Ephemeroptera	1	2.17	3341	7250
Hemiptera	4	6.25	99,144	619,650
Hymenoptera	10	3.43	118,308	405,796
Mecoptera	1	1.52	684	1040
Orthoptera	2	3.00	29,130	87,390
Thysanoptera	1	5.50	6521	35,866
Other orders		3.27	250,966	820,659
All insects		3.11	950,824	2,960,618

Notes: Studies that obtained molecular data for 10 or fewer individuals per morphology-based species (average among morphology-based species) were excluded (Table 1). For the 2 studies of the hemipteran genus *Bemisia*, we used the estimate from the study with >10 individuals sampled per morphology-based species. For the 2 studies of the hymenopteran genus *Cataglyphis*, we averaged the 2 estimates to get a mean ratio (1.13). For species numbers, we used the Catalogue of Life (Bánki et al. 2021), which was checked 14 March 2022 (number for each order in Supplementary Table S7). To project the number of species in each sampled order, we multiplied the ratio of molecular-based:morphology-based species by the number of described species. For all other orders, we used the mean ratio across the 8 sampled orders. We then multiplied this mean ratio by the summed number of species in these unsampled orders.

order and in insects in general (we assume that the vast majority of formally described insect species are morphology based).

Our sampling included only some insect orders (especially given that many orders have relatively few species). To accommodate the unsampled orders, we estimated an overall ratio across insects and assumed that the unsampled orders fit this general pattern. Specifically, we estimated the mean ratio of molecular to morphology-based species among the sampled orders, and then multiplied this mean ratio by the total number of described species among all the unsampled orders. The sampled orders spanned most described insect species (see Results), which suggests that the treatment of these unsampled orders should not strongly affect our overall estimates. We then summed the projected numbers of insect species across sampled and unsampled orders to obtain the total number of estimated molecular-based species among the described insect species.

We used the estimated number of molecular-based species divided by the number of described species to obtain an overall ratio for insects in general (Table 2). We then multiplied this ratio by projections of total insect richness (described and undescribed species) that were calculated using only morphology-based species. We specifically used the relatively recent projection of 6.8 million terrestrial arthropod species, consisting mostly of insects (Stork et al. 2015). This does not include insect-associated mites. As noted above, other projections of overall insect diversity are similar in magnitude, close to 6 million (e.g., Gaston 1991; Novotny et al. 2002; Basset et al. 2012). We assume that the current patterns of relative richness among insect orders will be maintained with 6 million or more morphology-based species, and we think that this assumption is consistent with these projections (e.g., Stork et al. 2015). Simply averaging among orders yields similar values to these richness-weighted projections (see Results), so violating this assumption need not strongly affect our conclusions.

We also estimated global species richness of other groups of organisms (Table 3). We followed Larsen et al. (2017) in assuming that species associated with insect hosts may be a major driver of global biodiversity in many groups. Following that review, we assumed that each insect species hosts (on average) one unique species of mite, nematode, microsporidian fungus, and apicomplexan protist. Based on the re-estimates of Wiens (2021), we assumed that each insect species hosts (on average) at last 7.6 unique species of bacteria (a value somewhat lower than the estimates of Larsen et al. 2017). We used the 4 basic scenarios of Larsen et al. (2017) regarding symbiont-associated diversity. These are described in Table 3.

When estimating global biodiversity, we assumed that animal diversity is dominated by insects and their symbionts (i.e., mites, nematodes). Animal species that are larger in body size than insects might host many more symbiont species. This should have little impact on overall species numbers given that there are far fewer projected species in these larger-bodied groups relative to described and projected numbers of insect species (e.g., Chapman 2009).

We performed these analyses of other groups on our primary estimates of cryptic insect diversity. We did not re-estimate global biodiversity for each of our supplementary analyses, since these generally gave similar estimates of overall insect diversity.

RESULTS

We focused first on papers that included nuclear genes and sampled >10 individuals per morphology-based species (on average). Using these criteria, we obtained usable data for 25 genera from 8 orders (Table 1; Fig. 2). These included many of the largest orders (Coleoptera, Diptera, Hemiptera, Hymenoptera), with the notable exception of Lepidoptera (but see below). The sampled orders spanned 73.4% of described insect

TABLE 3. Estimating global biodiversity of major groups of organisms based on the new estimates of cryptic insect diversity.

	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	Million species	% of total	Million species	% of total	Million species	% of total	Million species	% of total
Animals	84.59	9.4	84.59	13.6	84.59	3.8	52.87	9.4
Plants	0.34	0	0.34	0.1	0.34	0	0.34	0.1
Fungi	86.99	9.7	86.99	14	86.99	3.9	55.27	9.8
Protists	84.59	9.4	84.59	13.6	84.59	3.8	52.87	9.4
Bacteria	642.88	71.5	363.75	58.6	1949.48	88.4	401.81	71.3
Total	899.39		620.26		2205.99		563.16	

Notes: Scenario 1 assumes that all animal species have a full set of bacterial, protist, and fungal endosymbionts, even if they are parasites, but that microsporidian fungi and apicomplexan protists have little or no host-specific bacterial richness. Scenario 2 assumes that symbionts have limited numbers of symbionts (i.e., nematodes have an average of only one host-specific bacterial species) and that microsporidians and apicomplexans have few or no bacterial species. Scenario 3 assumes that all animal species have a full set of symbiont species and that microsporidians and apicomplexans host (on average) as many bacterial species as animal species do. Scenario 4 is identical to Scenario 1, except that it assumes that host-associated mites have more limited richness relative to their arthropod hosts (0.25 mites:other arthropod species). Archaean species richness is considered to be *relatively* small overall (Larsen et al. 2017), and so is not treated separately here. We assumed that each insect species hosts (on average) one mite species, and that each insect and mite species hosts (on average) one nematode species, yielding 84.59 million animal species for Scenarios 1–3. Plant richness is considered to be 0.34 million species. For fungi, we considered each animal species to host one microsporidian fungus species (84.59 million), and we added 2.4 million other fungal species (Larsen et al. 2017). For protists, we assumed that each animal species hosts at least one apicomplexan protist (and that this number renders other components of protist richness negligible in comparison). For bacteria, we assumed that each animal species hosts (on average) 7.6 bacterial species (Wiens 2021), and that (for Scenario 1) bacterial richness in plants, endosymbiotic fungi, and protists was relatively negligible. For Scenario 2, we assumed that each nematode species host only 1 unique bacterial species (on average) instead of 7.6. This yields a total 363.746 million bacterial species (42.296 million insect and mite species x 7.6 bacterial species and 42.296 million nematode species x 1 bacterial species). For Scenario 3, we assumed that all groups had the same mean number of host-specific bacteria as animals (7.6), including fungi and protists. Finally, for Scenario 4, we assumed a reduced number of mites relative to insects (0.25 mite species per insect host), but with other assumptions the same as Scenario 1. Thus, we assumed 26.435 million insect and mite species, for a total of 52.87 million animal species (including nematodes). For this scenario, we assumed that fungi and protists had negligible bacterial richness.

species. We calculated a mean ratio of molecular-based to morphology-based species among sampled genera in each order (Table 2). Among the 8 orders (Fig. 2), 6 had at least 2 cryptic species per morphology-based species (average across genera), and all had at least 1.5 (Table 2). The highest ratio of cryptic species was in Hemiptera, with 6.25 cryptic species per morphology-based species (Fig. 2; Table 2).

We multiplied the mean estimated ratio of cryptic (molecular-based) species for each order by the number of described species in that order (Table 2). For unsampled orders (26.6% of described insect species), we used the mean ratio among the 8 sampled orders (3.27). When we tallied up the projected number of species across all orders and divided this by the number of described species across all insects, the overall ratio of molecular-based to morphology-based species was 3.11.

We then used this latter number (3.11) to estimate the total number of insect species and overall patterns of biodiversity among non-insect groups. Multiplying this value (3.11) by a projected estimate of terrestrial arthropod diversity (6.8 million) yielded 21.15 million species. We then estimated richness across other major groups of organisms, using 4 basic scenarios for symbiont-associated diversity (Table 3). These scenarios yielded from 563.16 million to 2.206 billion species on Earth, of which 3.8–13.6% are animals, and 58.6–88.4% are bacteria.

We also estimated ratios of molecular to morphology-based species from our limited sample of studies using mitochondrial data only (Tables S8–S10). These estimates were broadly similar to those incorporating nuclear data. After excluding studies with limited
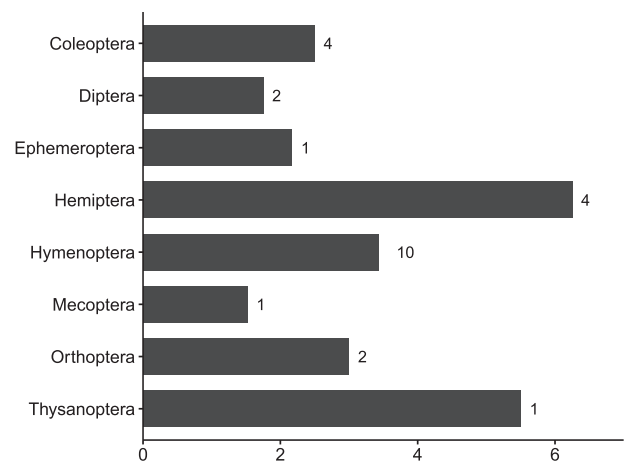

FIGURE 2. The ratio of estimated molecular-based species to morphology-based species, averaged among genera within each order. The number of genera sampled for each order is given on the right side of each bar. Data for each genus are given in Table 1 and are summarized for each order in Table 2.

sample sizes (Supplementary Table S9), we obtained usable data from 8 genera from 5 orders (Coleoptera, Diptera, Hemiptera, Lepidoptera, Orthoptera; Supplementary Table S10). The mean ratio of molecular-based to morphology-based species across these orders was 3.76 (Supplementary Table S10), similar to the mean across orders for nuclear data (3.27). Our set of usable studies based on nuclear data did not include any representatives of Lepidoptera (Table 2). The mitochondrial results for Lepidoptera (mean = 2.53 molecular-based species per morphology-based species) were

similar to the mean value across orders from nuclear data (3.27; Table 2).

We also evaluated the impact of our methodological choices on the primary analyses. First, we evaluated whether using the minimum or maximum number of species estimated in each study impacted the overall results (Tables S11–S16) relative to using the mean number. We found that they gave estimates of the overall ratio of cryptic insect species broadly similar to that for the mean (mean = 3.11, minimum = 2.53, maximum = 3.61). Second, we evaluated whether different minimum sample sizes (mean individuals sampled per morphology-based species) strongly changed the overall number of insect species estimated. We found that if we included studies with an average of 6 or more individuals per species (Supplementary Table S17), we included more studies (32 genera, 9 orders), and the overall ratio was similar but smaller (ratio = 2.54). Conversely, if we excluded studies with an average of <20 individuals per species (Supplementary Table S18), we included fewer studies (18 genera, 6 orders) and the overall ratio was similar but larger (ratio = 3.95).

We also tested whether there was a latitudinal gradient in the ratio of cryptic species within genera. The data for each genus are in Supplementary Table S19. We found a trend towards more cryptic species at lower latitudes (Fig. 3), but this was not significant ($r^2 = 0.117$; $P = 0.095$). We also found no consistent trend within orders. For example, in Coleoptera ($n = 4$ genera), the mean ratio was higher in temperate genera (mean = 3.75, $n = 2$) than tropical genera (mean = 1.25, $n = 2$), as in Orthoptera ($n = 2$ genera, temperate = 4.33, tropical = 1.67). In Hemiptera ($n = 4$ genera), the mean was instead somewhat higher in tropical genera (mean = 7.50, $n = 2$) than temperate genera (mean = 5, $n = 2$), and was much higher in Hymenoptera ($n = 10$; temperate mean = 1.95, $n = 6$; tropical mean = 5.75, $n = 4$).

Discussion

The number of species on Earth is highly uncertain, and insects are a pivotal group for estimating this number. Here, we estimated the average number of molecular-based cryptic species per morphology-based insect species (3.11; Table 2). We then extrapolated this number to estimate the total number of insect species based on the number of described insect species (0.95 million species described, yielding 2.96 million species total) and based on relatively well-established projections of undescribed insect richness from morphology (6.8 million total projected, yielding 21.15 million species). We also estimated overall global biodiversity across all groups, given that many species of mites, nematodes, fungi, protists, and bacteria may be associated with insect hosts. These projections yielded 0.56–2.20 billion species, the majority of which are bacteria (Table 3).
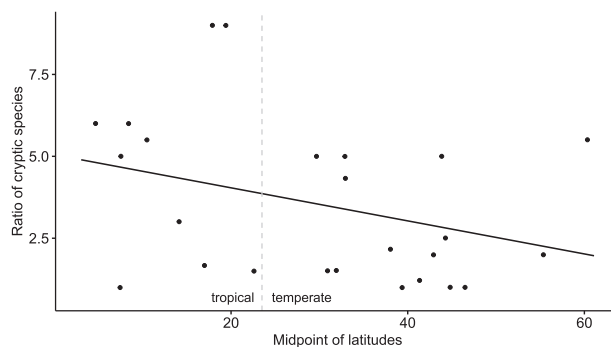

FIGURE 3. The relationship between the latitudinal position of each genus (mean among species) and the ratio of estimated molecular-based species to morphology-based species within that genus. Data for each genus are given in Supplementary Table S19. The vertical dashed line indicating tropical versus temperate latitudes at 23.4 degrees is somewhat arbitrary, but does not impact the analysis.

Comparison to Previous Studies

Our estimates of overall global biodiversity are broadly similar to those of Larsen et al. (2017), but with 2 important differences. First, those authors estimated that there are from 0.209 to 5.8 billion species on Earth (again with the majority being bacteria). They considered a range of possible ratios of molecular-based (cryptic) to morphology-based insect species, from 0 to 6. By providing improved estimates of cryptic insect diversity, we were able to narrow the overall range of numbers for global biodiversity, from roughly 30-fold to roughly 4-fold. That remaining 4-fold range hinges largely on different assumptions about how many symbionts the symbionts themselves host (e.g., how many bacteria does a protist or unicellular fungus host?).

Second, our estimate of the number of cryptic insect species is roughly half that of Larsen et al. (2017). Specifically, we estimated that there are approximately 3.11 cryptic species per morphology-based species, whereas those authors estimated that there were 5.9. What explains this 2-fold difference? Our sample size of insect genera is more than 3 times larger than that of Larsen et al. (2017). Perhaps most importantly, half of the genera that they sampled belonged to Hemiptera and Thysanoptera. Our study reveals that these 2 orders have ratios of molecular-based to morphology-based species that are roughly twice the estimated average number across orders (Fig. 2; Table 2). Our larger sampling here shows that the high numbers for those 2 orders are not broadly representative of insects overall.

Our overall estimates for animal richness and global biodiversity are strikingly larger than those of the excellent and well-cited study by Mora et al. (2011). There are 2 key differences between their estimates and ours. First, they did not consider molecular-based cryptic species. This explains why our estimates for animals are more than 3 times as large. Second, they projected total bacterial species richness that was roughly equivalent to described bacterial richness (~10,000 species). Many insect species seem to host numerous unique

bacterial species (almost all of which are undescribed). This inference is based on studies of bacterial species in closely related insects (which were used to obtain the estimate used here of 7.6 unique bacterial species per insect host species; Wiens 2021) and large-scale surveys across insects. For example, surveys of even modest numbers of insect host species (13–31) each found that these insects collectively host >1000 bacterial species (review in Wiens 2021). A study of 218 insect species across diverse orders found 9301 bacterial species (Yun et al. 2014). In summary, we think that assuming zero cryptic insect species and only ~10,000 bacterial species is no longer tenable. Our estimates of cryptic insect diversity and bacterial richness are doubtless imperfect, but using the number of described species instead seems very difficult to justify.

There are many other estimates of global biodiversity (e.g., Chapman 2009; Costello et al., 2012), and many studies of particular groups that are highly relevant (e.g., Novotny et al. 2002; Stork et al. 2015; Locey and Lennon 2016). However, few studies simultaneously incorporated: (a) all major groups of organisms, (b) cryptic species inferred from molecular data, and (c) host-associated species richness. We think that these 3 elements should become standard in estimates of global biodiversity going forward.

Our study also provides overall estimates of cryptic diversity across one of the largest clades of animals (based on described species). We know of few comparable estimates for other groups. A recent study in mammals estimated that there were ~0.66 cryptic species per described species (Parsons et al. 2022), but was based on mitochondrial DNA data alone. More broadly, it has been suggested that there are no significant differences in the relative number of cryptic species among clades and biogeographic regions (Pfenninger and Markus 2007). Our results do suggest that there might be such differences among insect orders (e.g., the high numbers in Hemiptera). We do not support a strong latitudinal increase in the ratio of cryptic species towards the equator, although one has been hypothesized (Freeman and Pennell 2021).

Potential Sources of Error

We acknowledge that there are various ways that our estimates of the numbers of cryptic insect species could be wrong. We think that the most important potential source of error is that species included in species-delimitation studies may be more likely to have cryptic species than the average, randomly chosen insect species. For example, the included species might have larger-than-average range sizes (and cryptic species typically occur in different parts of a morphology-based species' range; Larsen et al. 2017). This is a difficult factor to correct for.

Yet, many of the studies that we sampled here included 2 or more closely related species in a genus (Table 1), and were not just studies focused on a single broadly distributed species. Therefore, this issue need not be problematic overall. The most unbiased sampling might be obtained by sampling within-species variation in every (morphology-based) species in each genus. This should be done for multiple genera in the largest insect orders. As far as we know, such data are not yet available. Nevertheless, as an approximation, we re-estimated cryptic insect diversity after eliminating those nuclear-based studies that only considered a single morphology-based species (i.e., those studies that might be most impacted by this type of biased sampling). This deletion reduced the number of included genera to 16, and yielded mean ratios of molecular-based to morphology-based species that were identical to those including all studies for most orders (Supplementary Table S20; compare with Table 2). The only ones that changed were Hemiptera (from 6.25 to 5) and Hymenoptera, which changed more dramatically (3.43 to 1.09). The mean ratio across orders was similar (3.27 for all studies vs. 2.82 for multi-species studies only), as was the overall estimated ratio weighted by species numbers in orders (3.11 vs. 2.57). In summary, these results suggest that our estimates are not simply explained by biases associated with including studies that each focused on a single, morphology-based species.

Another possible solution is to consider other sampling strategies, such as focusing on species at a given site or sites rather than across species ranges. Studies of this type in insects have often found ~2 cryptic species per morphology-based species, including studies of Ephemeroptera, Plecoptera, and Trichoptera in Colorado and Ecuador (Polato et al. 2018), parasitoid wasp species at a site in Costa Rica (Smith et al. 2008), and cerambycid beetles at a site in French Guyana (Berkov 2002). These are simply examples, and not the result of a systematic survey. Nevertheless, they highlight that finding cryptic insect species is not necessarily contingent on broad-scale phylogeographic studies. We also note that studies based on single sites (or small sets of sites) will likely to underestimate cryptic diversity relative to studies that broadly sample across the geographic range of each species.

Another important way that our estimates could be wrong is that some methods or types of data might give biased answers. The best method(s) for estimating species limits from molecular data remain uncertain (e.g., Camargo et al. 2012; Rannala 2015; Sukumaran and Knowles 2017; Leaché et al. 2019; Chan et al. 2022). Most studies included in our primary analyses utilized nuclear data and 2 or more methods for species delimitation (70.4% of 27 studies in Table 1; mean = 2.5 methods per study overall). We primarily focused on results using the mean or plurality among studies. Yet, we obtained similar overall results using the maximum or minimum estimate of the number of species from each study (i.e., ~3 cryptic species per morphology-based species). Thus, our overall results do not appear to be particularly sensitive to variation among the methods used. Similarly, we obtained similar overall results using studies that included only mitochondrial data.

We also found that our conclusions were broadly robust to different minimum sample sizes of individuals for molecular data. However, it would be impossible to infer large numbers of cryptic species (which clearly do exist in some species) without sampling many individuals and localities.

Our estimates are also dependent on the sampling of studies in each insect group. There may be studies that our systematic search missed, and other relevant studies may have been published subsequently. Given the many downstream analyses, we did not constantly update our results with searches for new studies. However, simply adding more studies need not strongly impact our main conclusions. In our present results, those orders with the most sampled genera (Coleoptera = 4, Hemiptera = 4, Hymenoptera = 10; Table 2) have substantial estimated ratios of cryptic species (2.5, 6.3, 3.4), with two that are close to the overall mean estimate across insects (~3; except the larger number in Hemiptera). This strongly suggests that this overall mean estimate is not an artifact of limited sampling of genera within orders. Moreover, given their larger sample sizes, including more genera in these 3 well-sampled orders is less likely to change their results. Furthermore, these 3 orders together encompass roughly half of all described insects (Table 2). Therefore, these 3 orders are pivotal for the overall estimate for insects. By contrast, those orders with smaller ratios of cryptic species (Diptera, Ephemeroptera, Mecoptera) have only 1 or 2 genera sampled (Fig. 2) and together encompass <20% of described insect species (Table 2). Given their limited sampling, mean estimates in these 3 orders are especially sensitive to sampling more genera (which could bring the estimate for each of these orders closer to the mean estimated ratio across all orders). Yet, if additional sampling of genera yielded similar numbers in these 3 orders, this would have little consequence for the overall estimate for insects given the limited collective richness of these 3 orders. In summary, these results suggest that additional sampling of insect genera is by itself unlikely to overturn our conclusions (all else being equal).

We acknowledge that some readers might dismiss our results simply because the overall number of genera sampled is small (25 genera; Table 1) relative to the total number of genera, families, and species. Yet, we obtained similar ratios of cryptic species after including 7 additional genera (with reduced sampling of individuals) and 8 genera with mitochondrial data. Furthermore, most orders showed similar patterns of cryptic richness, with some genera having few or no cryptic species and others having many molecular-based species per morphology-based species (range of ratios among orders: Coleoptera: 1–5.5; Diptera: 1–2.5; Hymenoptera: 1–9; Orthoptera: 1.67–4.33; Table 2). Based on mitochondrial data, Lepidoptera also fits this pattern (range: 1.04–4.00). Four of 8 orders also have mean ratios between 2 and 4 (Table 2). We think that sampling every family (or every order) is not as important for estimating insect-wide richness as estimating these overall patterns in the most species-rich orders. We note that Hemiptera is somewhat exceptional in having consistently high cryptic richness. Various ways of reducing our sample size of genera also yielded similar estimated ratios, either by eliminating genera with <20 individuals sampled (n = 18 genera included, estimated ratio = 3.95; Supplementary Table S18), deleting those with only one sampled species (n = 16 genera included, estimated ratio = 2.57; Supplementary Table S20), and including only genera with mitochondrial data (n = 8 genera, estimated ratio = 3.76; Supplementary Table S10). Overall, we show that an estimated ratio close to 3 is robust to various additions and deletions of sampled genera.

There are numerous other potential factors that might affect whether a particular morphology-based species has cryptic species, and how many. These include dispersal ability, patterns of morphological evolution, and taxonomic practices within a given order. Yet, the mean values of cryptic species ratios for orders varied along a limited range of values (1.5–6.2), with 4 of 8 sampled orders having from 2.2 to 3.4. Thus, even if there are biases, the overall variability in these mean numbers seemed limited. We also did not find a strong impact of the geographic location of genera (i.e., temperate vs. tropical), although this might be worth exploring further, especially within orders (like Hymenoptera). If there are indeed more cryptic species in the tropics, this only reinforces our main conclusion that cryptic insect diversity is important and must be included in future estimates of global biodiversity.

Putting aside the question of cryptic species, our overall estimates of global biodiversity could also be wrong, especially if the estimates of Larsen et al. (2017) for insect-associated groups are grossly incorrect. Specifically, their estimates for the mean numbers of mites, nematodes, apicomplexan protists, microsporidian fungi, and bacteria hosted by each insect species might be wrong. Some of this uncertainty is addressed in Table 3. To better address this uncertainty, systematic studies of the number of host-associated species of these groups among closely related insect species from several major insect orders are needed. However, it is important to note that the estimates of Larsen et al. (2017) could be overestimates or underestimates for each of these groups. Simply assuming that there are few or no species of these 5 groups associated with insect hosts seems empirically unsupported.

We also think that there are potential issues that are not actually problematic. First, the number of morphology-based species that are synonyms of other morphology-based species seems unlikely to overwhelm the number of cryptic species (e.g., Stork et al. 2015). Species-delimitation analyses based on molecular data can infer that morphology-based species are not actually distinct. Therefore, our results already take the possibility of synonyms into account, and show that morphology-based synonyms do not counterbalance the number of morphologically cryptic species. Second,

some readers might dismiss morphologically cryptic species altogether. We strongly reject this idea. We believe that species are real entities in nature that do not rely on morphological diagnosis for their existence. Therefore, we see no reason why the only species that are real are those that can be diagnosed morphologically by humans. Finally, some readers may be concerned about when and how all these cryptic species will be described. We share this concern. However, the idea that we should ignore cryptic species because it will be difficult to describe them all seems like the tail wagging the dog. We also note that there have been calls to use DNA data to accelerate the pace of species discovery (e.g., Tautz et al. 2003), but these calls have also met with considerable resistance (e.g., Lipscomb et al. 2003). Some of this resistance may be confounded with the practice of DNA barcoding with a single mitochondrial gene.

Conclusions

In this study, we have provided improved estimates of the number of cryptic insect species hidden in morphology-based species. This is not some obscure and trivial topic. We show that these estimates of cryptic insect diversity are vital to estimating global biodiversity. Our study greatly improves on the previous estimates of cryptic insect diversity, more than tripling the number of included genera and cutting the estimated number of cryptic species in half. Furthermore, our results provide a relatively narrow range of global biodiversity estimates. Rather than ranging from the low millions to the trillions, our estimates span a 4-fold range centered near ~1 billion species. Our estimates here will not be the last word on cryptic insect species (or global biodiversity). Instead estimates of cryptic diversity should improve over time as more case studies are published and the data and methods for species delimitation are expanded and refined. We think that the worst approach is to do what is often done now: to simply ignore cryptic insect species diversity when estimating global biodiversity. Even though estimates of cryptic diversity may change over time, there is clearly no justification for assuming that the number of cryptic species is zero.

Supplementary Materials

Supplementary material, including data files and online-only appendices, can be found in the Dryad data repository a http://datadryad.org, http://dx.doi.org/10.5061/dryad.vt4b8gtt7.

Data Availability

The data underlying this article are available in its online supplementary material (specifically, Tables S1–S20 and Datasets S1–S2).

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? Syst. Biol. 63:518–533.

Aguilar-Velasco R.G., Poteaux C., Meza-Lazaro R., Lachaud J.P., Dubovikoff D., Zaldivar-Riveron A. 2016. Uncovering species boundaries in the Neotropical ant complex *Ectatomma ruidum* (Ectatomminae) under the presence of nuclear mitochondrial paralogues. Zool. J. Linn. Soc. 178:226–240.

Ahrens D., Fabrizi S., Sipek P., Lago P.K. 2013. Integrative analysis of DNA phylogeography and morphology of the European rose chafer (*Cetonia aurata*) to infer species taxonomy and patterns of postglacial colonisation in Europe. Mol. Phylogenet. Evol. 69:83–94.

Bandelt H.J., Forster P., Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16:37–48.

Bánki O., Roskov Y., Vandepitte L., DeWalt R.E., Remsen D., Schalk P., Orrell T., Keping M., Miller J., Aalbu R., Adlard R., Adriaenssens E., Aedo C., Aescht E., Akkari N., Alonso-Zarazaga M.A., Alvarez B., Alvarez F., Anderson G., et al. 2021. Catalogue of life checklist (Version 2021-08-25). Catalogue of Life. checked 4 September 2021. Available from: http://www.catalogueoflife.org.

Basset Y., Cizek L., Cuénoud P., Didham R.K., Guilhaumon F., Missa O., Novotny V., Ødegaard F., Roslin T., Schmidl J., Tishechkin A.K., Winchester N.N., Roubik D.W., Aberlenc H., Bail J., Barrios H., Bridle J.R., Castaño-Meneses G., Corbara B., Curletti G., Duarte da Rocha W., De Bakker D., Delabie J.H.C., Dejean A., Fagan L.L., Floren A., Kitching R.L., Medianero E., Miller S.E., Gama de Oliveira E., Orivel J., Pollet M., Rapp M., Ribeiro S.P., Roisin Y., Schmidt J.B., Sørensen L., Leponce M. 2012. Arthropod diversity in a tropical forest. Science 338:1481–1484.

Berkov A. 2002. The impact of redefined species limits in *Palame* (Coleoptera: Cerambycidae: Lamiinae: Acanthocinini) on assessments of host, seasonal, and stratum specificity. Biol. J. Linn. Soc. 76:195–209.

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 Ecol. Evol. 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.

Chamberlain S.A., Boettiger C. 2017. R Python, and Ruby clients for GBIF species occurrence data. PeerJ Preprints. 5:e3304v1. doi:10.7287/peerj.preprints.3304v1

Chan K.O., Hutter C.R., Wood P.L., Su Y.C., Brown R.M. 2022. Gene flow increases phylogenetic structure and inflates cryptic species

estimations: a case study on widespread Philippine puddle frogs (*Occidozyga laevis*). Syst. Biol. 71:40–57.

Chapman A.D. 2009. Numbers of living species in Australia and the world. 2nd ed. Toowoomba (Australia): Australian Biodiversity Information Services.

Chen Y.T., Tseng H.Y., Jeng M.L., Su Y.C., Huang W.S., Lin C.P. 2017. Integrated species delimitation and conservation implications of an endangered weevil *Pachyrhynchus sonani* (Coleoptera: Curculionidae) in Green and Orchid Islands of Taiwan. Syst. Entomol. 42:796–813.

Cook L.G., Rowell D.M. 2007. Genetic diversity, host-specificity and unusual phylogeography of a cryptic, host-associated species complex of gall-inducing scale insects. Ecol. Entomol. 32:506–515.

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

Darwell C.T., al-Beidh S., Cook J.M. 2014. Molecular species delimitation of a symbiotic fig-pollinating wasp species complex reveals extreme deviation from reciprocal partner specificity. BMC Evol. Biol. 14:189.

de Moya R.S., Brown J.K., Sweet A.D., Walden K.K.O., Paredes-Montero J.R., Waterhouse R.M., Johnson K.P. 2019. Nuclear orthologs derived from whole genome sequencing indicate cryptic diversity in the *Bemisia tabaci* (Insecta: Aleyrodidae) complex of whiteflies. Diversity 11:151.

Delgado-Machuca N., Meza-Lázaro R.N., Romero-Nápoles J., Sarmiento-Monroy C.E., Amarillo-Suárez A.R., Bayona-Vásquez N.J., Alejandro Z.R. 2020. Genetic structure, species limits and evolution of the parasitoid wasp genus *Stenocorse* (Braconidae: Doryctinae) based on nuclear 3RAD and mitochondrial data. Syst. Entomol. 45:33–47.

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

Eyer P.A., Hefetz A. 2018. Cytonuclear incongruences hamper species delimitation in the socially polymorphic desert ants of the *Cataglyphis albican*s group in Israel. J. Evol. Biol. 31:1828–1842.

Eyer P.A., Seltzer R., Reiner-Brodetzki T., Hefetz A. 2017. An integrative approach to untangling species delimitation in the *Cataglyphis bicolor* desert ant complex in Israel. Mol. Phylogenet. Evol. 115:128–139.

Falush D., Stephens M., Pritchard J.K. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587.

Fossen E., Ekrem T., Nilsson A., Bergsten J. 2016. Species delimitation in northern European water scavenger beetles of the genus *Hydrobius* (Coleoptera, Hydrophilidae). Zookeys 564:71–120.

Freeman B.G., Pennell M.W. 2021. The latitudinal taxonomy gradient. Trends Ecol. Evol. 36:778–786.

Gaston K.J. 1991. The magnitude of global insect species richness. Conserv. Biol. 5:283–296.

GBIF.org. 2022. GBIF occurrence download. doi:10.15468/dl.nkxq9w. Available from: https://github.com/ropensci/rgbif

Grimaldi D., Engel M.S. 2005. Evolution of the insects. Cambridge (UK): Cambridge University Press.

Groombridge B., Jenkins M.D. 2002. World atlas of biodiversity. Berkeley (CA): University of California Press.

Gueuning M., Frey J.E., Praz C. 2020. Ultraconserved yet informative for species delimitation: ultraconserved elements resolve long-standing systematic enigma in Central European bees. Mol. Ecol. 29:4203–4220.

Gwiazdowski R.A., Vea I.M., Andersen J.C., Normark B.B. 2011. Discovery of cryptic species among North American pine-feeding *Chionaspis* scale insects (Hemiptera: Diaspididae). Biol. J. Linn. Soc. 104:47–62.

Hammond P. M. 1995. The current magnitude of biodiversity. In: Heywood V. H., editor. Global biodiversity assessment. Cambridge (UK): Cambridge University Press. p. 113–128

Hernández-López A., Rougerie R., Augustin S., Lees D.C., Tomov R., Kenis M., Çota E., Kullaj E., Hansson C., Grabenweger G., Roques A., López-Vaamonde C. 2012. Host tracking or cryptic adaptation?

Phylogeography of *Pediobius saulius* (Hymenoptera, Eulophidae), a parasitoid of the highly invasive horse-chestnut leafminer. Evol. Appl. 5:256–269.

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. Mol. Phylogenet. Evol. 76:172–180.

Hu G.L., Gao K., Wang J.S., Hebert P.D.N., Hua B.Z. 2019. Molecular phylogeny and species delimitation of the genus *Dicerapanorpa* (Mecoptera: Panorpidae). Zool. J. Linn. Soc. 187:1173–1195.

Jin Q., Hu X.M., Han H.L., Chen F., Cai W.J., Ruan Q.Q., Liu B., Luo G.J., Wang H., Liu X., Ward R.D., Wu C.S., Wilson J.J., Zhang A.B. 2018. A two-step DNA barcoding approach for delimiting moth species: moths of Dongling Mountain (Beijing, China) as a case study. Sci. Rep. 8:14256.

Jusoh W.F.A., Ballantyne L., Chan K.O. 2020. DNA-based species delimitation reveals cryptic and incipient species in synchronous flashing fireflies (Coleoptera: Lampyridae) of Southeast Asia. Biol. J. Linn. Soc. 130:520–532.

Larsen B.B., Miller E.C., Rhodes M.K., Wiens J.J. 2017. Inordinate fondness multiplied and redistributed: the number of species on Earth and the new Pie of Life. Q. Rev. Biol. 92:229–265.

Leaché A.D., Zhu T., Rannala B., Yang Z. 2019. The spectre of too many species. Syst. Biol. 68:168–181.

Lin X., Stur E., Ekrem T. 2018. Exploring species boundaries with multiple genetic loci using empirical data from non-biting midges. Zool. Scr. 47:325–341.

Lin Y.P., Edwards R.D., Kondo T., Semple T.L., Cook L.G. 2017. Species delimitation in asexual insects of economic importance: the case of black scale (*Parasaissetia nigra*), a cosmopolitan parthenogenetic pest scale insect. PLoS One 12:e0175889.

Lipscomb D., Platnick N., Wheeler Q. 2003. The intellectual content of taxonomy: a comment on DNA taxonomy. Trends Ecol. Evol. 18:65–66.

Locey K.J., Lennon J.T. 2016. Scaling laws predict global microbial diversity. Proc. Natl. Acad. Sci. USA 113:5970–5975.

Martinet B., Lecocq T., Brasero N., Biella P., Urbanová K., Valterová I., Cornalba M., Gjershaug J.O., Michez D., Rasmont P. 2018. Following the cold: geographical differentiation between interglacial refugia and speciation in the arcto-alpine species complex *Bombus monticola* (Hymenoptera: Apidae). Syst. Entomol. 43:200–217.

Matenaar D., Fingerle M., Heym E., Wirtz S., Hochkirch A. 2018. Phylogeography of the endemic grasshopper genus *Betiscoides* (Lentulidae) in the South African Cape Floristic Region. Mol. Phylogenet. Evol. 118:318–329.

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 Biol. 9:e1001127.

Novotny V., Bassett 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.

Parsons D.J., Pelletier T.A., Wieringa J.G., Duckett D.J., Carstens B.C. 2022. Analysis of biodiversity data suggests that mammals species are hidden in predictable places. Proc. Natl. Acad. Sci. USA 119:e2103400119.

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). Mol. Phylogenet. Evol. 84:220–231.

Pfenninger M., Schwenk K. 2007. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. BMC Evol. Biol. 7:1211–1216.

Polato N.R., Gill B.A., Shah A.A., Gray M.M., Casner K.L., Barthelet A., Messer P.W., Simmons M.P., Guayasamin J.M., Encalada A.C., Kondratieff B.C., Flecker A.S., Thomas S.A., Ghalambor C.K., Poff N.L., Funk W.C., Zamudio K.R. 2018. Narrow thermal tolerance and low dispersal drive higher speciation in tropical mountains. Proc. Natl. Acad. Sci. USA 115:12471–12476.

Popović D., Acanski J., Djan M., Obreht D., Vujic A., Radenkovic S. 2015. Sibling species delimitation and nomenclature of the *Merodon avidus* complex (Diptera: Syrphidae). Eur. J. Entomol. 112:790–809.

Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Rannala B. 2015. The art and science of species delimitation. Curr. Zool. 61:846–853.

Raven P.H., Yeates D.K. 2007. Australian biodiversity: threats for the present, opportunities for the future. Aust. J. Entomol. 46:177–187.

Ross K.G., Gotzek D., Ascunce M.S., Shoemaker D.D. 2010. Species delimitation: a case study in a problematic ant taxon. Syst. Biol. 59:162–184.

Rutschmann S., Detering H., Simon S., Funk D.H., Gattolliat J.L., Hughes S.J., Raposeiro P.M., DeSalle R., Sartori M., Monaghan M.T. 2017. Colonization and diversification of aquatic insects on three Macaronesian archipelagos using 59 nuclear loci derived from a draft genome. Mol. Phylogenet. Evol. 107:27–38.

Singhal S., Hoskin C.J., Couper P., Potter S., Moritz C. 2018. A framework for resolving cryptic species: a case study from the lizards of the Australian wet tropics. Syst. Biol. 67:1061–1075.

Smith M.A., Rodriguez J.J., Whitfield J.B., Deans A.R., Janzen D.H., Hallwachs W., Hebert P.D. 2008. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proc. Natl. Acad. Sci. USA 105:12359–12364.

Stork N.E., McBroom J., Gely C., Hamilton A.J. 2015. New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. Proc. Natl. Acad. Sci. USA 112:7519–7523.

Struck T.H., Feder J.L., Bendiksby M., Birkeland S., Cerca J., Gusarov V.I., Kistenich S., Larsson K.H., Liow L.H., Nowak M.D., Stedje B., Bachmann L., Dimitrov D. 2018. Finding evolutionary processes hidden in cryptic species. Trends Ecol. Evol. 33:153–163.

Sukumaran J., Knowles L.L., Knowles L.L. 2017. Multispecies coalescent delimits structure, not species. Proc. Natl. Acad. Sci. USA 114:1607–1612.

Tautz D., Arctander P., Minelli A., Thomas R.H., Vogler A.P. 2003. A plea for DNA taxonomy. Trends Ecol. Evol. 18:70–74.

Tollefson J. 2019. Humans are driving one million species to extinction. Nature 569:171.

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. New York: Springer.

Wiens J.J. 2021. Vast (but avoidable) underestimation of global biodiversity. PLoS Biol. 19:e3001192.

Yang Z., Rannala B., Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. USA 107:9264–9269.

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., Drake H.L. 2014. Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage and phylogeny of host. Appl. Environ. Microbiol. 80:5254–5264.

Zhu X.C., Jing C., Rui C., Jiang L.Y., Qiao G.X. 2017. DNA barcoding and species delimitation of Chaitophorinae (Hemiptera, Aphididae). ZooKeys 656:25–50.