

Inferring global species richness from megatranssect data and undetected species estimates

Bernhard A. Huber

Alexander Koenig Research Museum of Zoology, Adenauerallee 160, 53113 Bonn, Germany
b.huber@leibniz-zfmk.de

Anne Chao

Institute of Statistics, National Tsing Hua University, Hsin-Chu, 30043, Taiwan

Abstract

Ratio-like approaches for estimating global species richness have been criticised for their unjustified extrapolation from regional to global patterns. Here we explore the use of cumulative percentages of ‘new’ (i.e., not formally described) species over large geographic areas (‘megatranssects’) as a means to overcome this problem. In addition, we take into account undetected species and illustrate these combined methods by applying them to a family of spiders (Pholcidae) that currently contains some 1,700 described species. The raw global cumulative percentage of new species (‘new’ as of the end of 2008, when 1,001 species were formally described) is 75.1%, and is relatively constant across large biogeographic regions. Undetected species are estimated using the *Chao2 estimator* based on species incidence data (date by species and locality by species matrices). The estimated percentage of new species based on the date by species matrices is 76.0% with an estimated standard error (s.e.) of 2.6%. This leads to an estimated global species richness of about 4,200 with a 95% confidence interval of (3,300, 5,000). The corresponding values based on locality by species matrices are 84.2% (s.e. 3.0%) and 6,300 with a 95% confidence interval of (4,000, 8,600). Our results suggest that the currently known 1,700 species of Pholcidae may represent no more than about 25–40% of the total species richness. The impact of further biasing factors like geography, species size and distribution, cryptic species, and model assumptions needs to be explored.

Keywords

species richness – estimation – *Chao2* estimator – Pholcidae – megatranssect

Introduction

Species are widely considered a fundamental unit in biology, and this explains to a large extent the continuous efforts to estimate species numbers, both globally and at various taxonomic and geographic levels (Caley et al., 2014; Larsen et al., 2017). Species richness is a key metric that is relevant far beyond purely academic purposes, with immense importance in evolution research, ecology, and conservation. A large variety of methods have been proposed to estimate species richness, ranging from expert opinion to relationships of body size and species number; rates of species description; latitudinal species richness gradients; ratios of known to unknown faunas; extrapolations from samples; species turnover; and sophisticated hierarchical multispecies models (Stork, 1997; Mora et al., 2011; Costello et al., 2012; Scheffers et al., 2012; Iknayan et al., 2014; Stork et al., 2015). Despite the long term efforts and the plethora of methods, however, global species richness estimates have not converged (Caley et al., 2014; but see Stork et al., 2015).

The majority of studies suggest that a relatively small part of the global eukaryotic species richness is formally described (Hamilton et al., 2010, 2011; Mora et al., 2011; Stork et al., 2015; but see Costello et al., 2012, 2013; Costello & Chaudhary, 2017). The fundamental difficulty is thus to estimate the unknown majority from a relatively limited data set. In principle, the estimates can be improved along two lines: by adding data, and by improving the analytical methods. The first approach is clearly more time-consuming, more expensive, and usually heavily dependent on expert collectors and taxon specialists. Yet it provides the essential data that determine the usefulness of any subsequent analysis. Here we combine both approaches; we use raw data derived from extensive recent collecting and verified according to current taxonomic

knowledge, and we analyse these data based on a robust statistical estimation method which takes into account undetected species.

Spiders currently contain some 48,000 nominal species (World Spider Catalog, 2019), and recent estimates of global spider species richness range from 76,000 to 170,000 (Coddington & Levi, 1991; Platnick, 1999; Agnarsson et al., 2013; Platnick & Raven, 2013). This range of estimates (1:2.2) is smaller than that of recent global species richness estimates (e.g., Caley et al., 2014: ~1:4.5), but it is still strong evidence for incomplete data and for extrapolations based on debatable assumptions. The present paper deals with spider species richness at a lower hierarchical level, taking as an example one of the currently known 117 spider families, the Pholcidae. Pholcidae are widely known by their synanthropic species but most of the family's diversity (currently some 1,700 species) occurs in tropical and subtropical regions around the globe (Huber, 2018). Here they have adapted to a wide range of microhabitats, which in turn may have contributed to the family's high species richness (Eberle et al., 2018). Except for a few species associated with humans, most species seem to have small ranges of distribution (e.g., Huber et al., 2014; Huber, 2015; Huber & Carvalho, 2019), and even at the level of genera, there is very little overlap among major geographic regions (e.g., only two of the currently known 86 genera have representatives both in the Old and New Worlds; Huber, 2011). More than 1,000 species have been described since 2000. The cumulative curve of currently valid pholcid species (fig. 1) and extensive material of scientifically unknown species available in collections suggest that we are still far from a complete taxonomic knowledge of the family. Here we explore the potential of a ratio approach to render this "far from complete" more precisely. We apply a novel method that combines field data gathered over large geographic areas ('megatranssects') with the

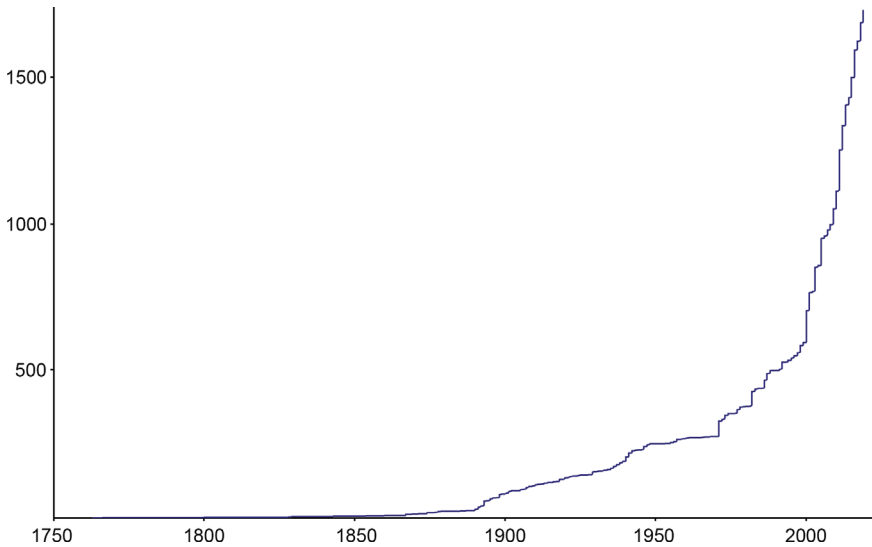


FIGURE 1 Cumulative curve of currently valid species of Pholcidae (y-axis) as a function of time (x-axis). The curve suggests that we are far from approaching a complete taxonomic knowledge of the family.

Chao2 undetected species estimator. Our data originate from extensive sampling efforts in South America, Africa, Southeast Asia, the Caribbean, and Subtropical and Mediterranean regions (a total of 31 trips, 449 field days, and 441 visited localities). The *Chao2* estimator is a statistical method to infer total species richness (and thus undetected species richness) in incomplete samples based on species incidence data in multiple sampling units. By combining these two methods we hope to overcome a major weakness of the ratio approach and of any approach ignoring undetected species.

Material and methods

Megatranssect collecting

The basic approach used here employs percentages of 'new' (i.e., not formally described) species collected during intensive sampling events (the *Known to unknown ratio* method in Mora et al., 2011). Under the assumption that the global ratio of known (described) species and unknown (undescribed/new) species is approximately the same as that of

a particular sample, we can infer the global number of unknown new species based on the currently known species richness provided such a ratio is available from the particular sample. This would be a simple method to extrapolate global species richness. The above assumption is equivalent to assuming that the percentage of new species in a particular sample is representative of the global percentage of new species. In contrast to previous studies using this approach (e.g., Hodgkinson & Casson, 1991; Grassle & Maciolek, 1992), we try to avoid two pitfalls: (1) by comparing results from megatranssects in different continents we reduce the potential impact of unjustified extrapolation from regional to global species richness; (2) by focusing on one particular family we can largely avoid synonyms and misidentifications and we avoid further assumptions needed for extrapolating to total spider or even total species richness.

Our raw data are based on the results of a series of 31 collecting trips to South America, the Caribbean, tropical Africa, Southeast Asia, and a few subtropical and Mediterranean regions, conducted between 2002 and 2017

(for details, see supplementary tables S1–S5). The primary aim at each of these expeditions was to collect a maximum of pholcid species present at each locality. A locality is here roughly defined as an area with a radius of approximately 5 km. By this definition, a total of 441 localities were sampled in 449 field days (table 1). Each location or each day can be treated as a sampling unit. Both kinds of sampling units are statistically meaningful in our estimation framework. If data are sufficient, a robust estimation method should give generally consistent results for the two kinds of sampling units.

In the course of sampling, common and ‘easy’ species were deliberately ignored after an initial period and the effort was increasingly focused on more cryptic and ‘difficult’ species. Collecting was done manually, with a variety of strategies adapted to different microhabitats. At each locality and for each microhabitat, the search was terminated when the subjective impression was formed that all species had been detected, regardless of the amount of time taken to reach that point. In each sampling unit (a location or a sampling date), only species detection/non-detection was recorded. For both practical and biological reasons, recording species detection/non-detection in multiple sampling units is often preferable to enumerating individuals in a single sampling unit (abundance data) so that an exact count of individuals is not necessary. From a statistical perspective, detection/non-detection data over multiple sampling units support statistical approaches to richness estimation that are just as powerful as corresponding abundance-based approaches (Chao et al., 2014). A further advantage is that replicated incidence records account for spatial (and temporal) heterogeneity in the data (Colwell et al., 2004, 2012).

Vouchers of all species from all localities are deposited at the Alexander Koenig Research Museum of Zoology, Bonn (ZFMK).

Dataset

Due to the long time span of data collecting, a threshold had to be set for classifying species as ‘known’ (formally described) or ‘new’ (not formally described). The end of 2008 was chosen because most species newly discovered during the 31 collecting trips were described after 2008 (or are still undescribed). In addition, the end of 2008 makes a convenient base line, with exactly 1,001 currently valid species known at the time (World Spider Catalog, 2019; Huber, 2018). Cumulative percentages of new species are used to visualize changes in the known-to-unknown ratio as a function of cumulative field days. This ratio, while expected to show wide initial variation due to chance events (e.g., the first field day may produce 100% new species or 0% new species) should stabilize over time. However, stabilization may not only result from similar percentages of new species collected at different localities but also from saturation, i.e. from a lack of any additional species not yet collected before, whether new or not. For this reason, absolute numbers of cumulative total and new species were also plotted against cumulative field days.

Instead of plotting global curves, separate curves were plotted for the major geographic regions. This is justified because synanthropic species (those associated with humans) were excluded from the analysis and this resulted in nearly zero species overlap among geographic regions (only one species was shared between Africa and the Mediterranean). In addition, this separation was intended to reveal differences among major geographic regions that would likely be lost in a global plot.

Undetected species estimation

Undetected species were estimated using the *Chao2 estimator* (Chao & Chiu, 2016) based on date by species and locality by species incidence matrices. The *Chao2 estimator* originally developed in Chao (1987) refers

TABLE 1 Main results divided by geographic regions, derived from supplementary tables S1–S5, and from incidence matrices in supplementary tables S6–S15

Region	Trips	Days	Localities	# Species		% New Species		
				Raw Total	Raw New	Ratio of Raw Numbers	<i>Chao2</i> : Date by Species	<i>Chao2</i> : Locality by Species
South America	9	132	101	242	180	74.4%	79.3%	84.9%
Africa	6	98	97	105	75	71.4%	78.4%	80.6%
Asia	6	114	110	225	181	80.4%	80.8%	84.7%
Caribbean	4	58	62	79	61	77.2%	84.8%	87.0%
Subtropical and Mediterranean	6	47§	71§	38	20	52.6%	56.6%	83.7%
Total	31	449	441	688*	517			
Average						71.2%	76.0%	84.2%

§ One field day and one locality are excluded in the analysis because no species (other than synanthropics) were found on that day in that locality.

* 687 species are non-shared species among regions; one species is shared between Africa and the Mediterranean; global raw cumulative percentage of new species = $517/688 = 75.1\%$.

to the species richness estimator (observed plus undetected) based on incidence (detection/non-detection) data for multiple sampling units. In our analysis, we consider two kinds of sampling units: a locality and a field sampling date. The intuitive concept behind the theory is that frequent species (which are detected in many sampling units) contain almost no information about the undetected species richness, whereas the detected rare/infrequent species (which are infrequently detected) contain almost all the information about the undetected species richness.

The *Chao2 estimator* of the number of undetected species is based on the count of “unique” species (denoted by Q_1 , i.e., the number of species that are detected in only one sampling unit), and the count of “duplicate” species (denoted by Q_2 , i.e., the number of species that are detected in only two sampling units). Chao (1987) derived a lower bound of the undetected species richness

as $(1 - 1/T) Q_1^2 / (2Q_2)$, where T denotes the number of sampling units. Adding the number of detected species to this estimate, we obtain the *Chao2 estimator* of species richness: $\hat{S}_{Chao2} = S_{obs} + (1 - 1/T) Q_1^2 / (2Q_2)$, where S_{obs} denotes the number of species detected in the data. Although these estimates of undetected and total species richness were derived as lower bounds, they are nearly unbiased under certain conditions (Chao & Colwell, 2017) (see Discussion). The method holds even if species are spatially aggregated or clustered in the study area. Chao (1987) also derived an analytical standard error to reflect sampling variation so that a confidence interval of species richness can be obtained.

Within each region, the *Chao2* estimate was calculated separately for described species data and for undescribed/new species data. The resulting *Chao2* estimates were then used to determine the percentages of described and undescribed species in each region. The

Chao2 estimator and the associated inference can be computed from the online software SpadeR (Species-richness Prediction and Diversity Estimation in R) available at <https://chao.shinyapps.io/SpadeR/>. All the numerical results are shown in supplementary tables S6–S15.

Results

In the three major megatransects (98–132 field days: South America, Africa, Southeast Asia; table 1), cumulative percentages of ‘new’ species (i.e., not formally described as of end of 2008) stabilize at about 74.4%, 71.4% and 80.4%, respectively (fig. 2). In each case, absolute

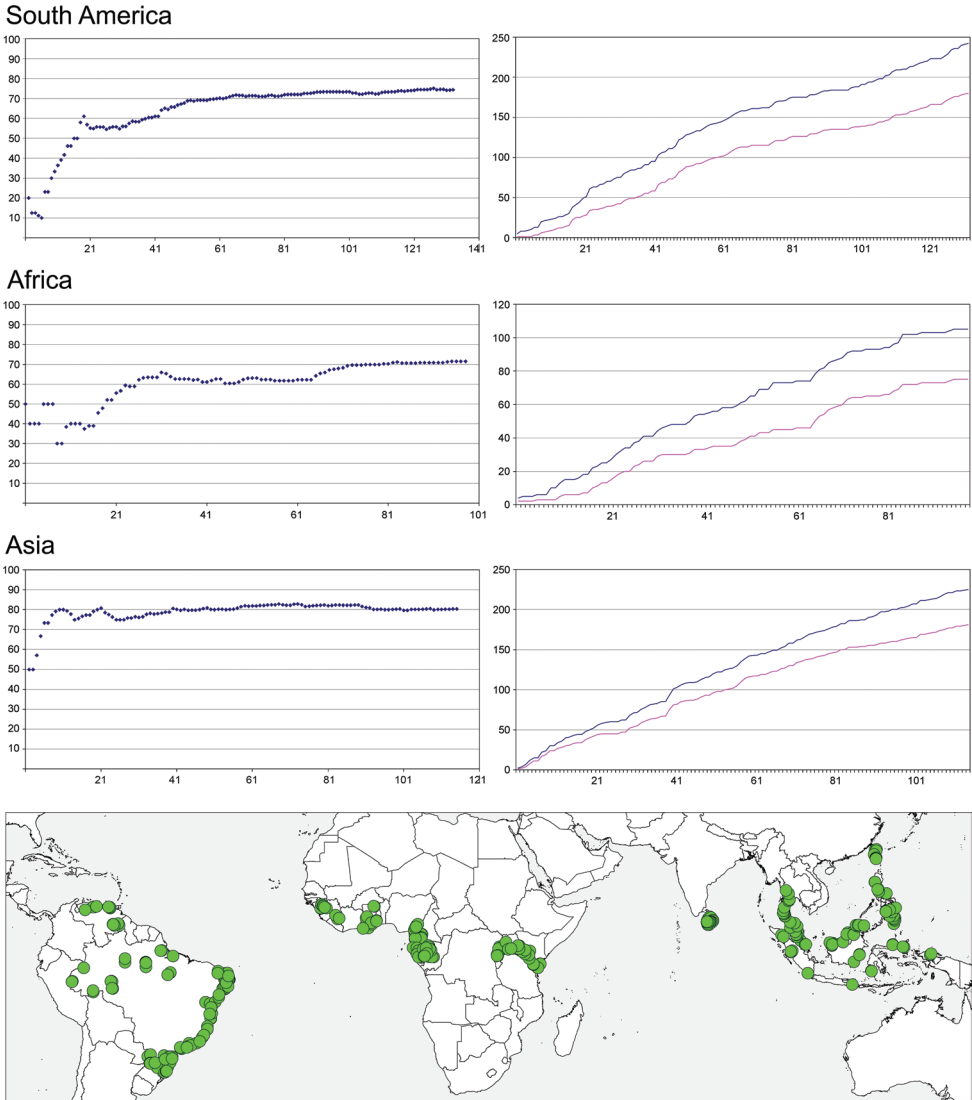


FIGURE 2 Cumulative percentages of new species (y-axis) as a function of cumulative field days (x-axis) (left) and cumulative number of total (upper/blue line) and new (lower/red line) species (y-axis) as a function of cumulative field days (x-axis) (right), for the three major tropical megatransects shown on the map. Each green dot represents a sampling locality. For raw data of all geographic regions, see supplementary tables S1–S5.

numbers of cumulative total and new species increase more or less constantly over time (fig. 2, right panels), showing that there are still many undetected species and that the stabilization of the cumulative percentages (left panels) is not the result of saturation. The smaller Caribbean transect (58 field days) ends with a similar percentage (77.2%; table 1) but does not seem to have stabilized (supplementary table S4). All subtropical and Mediterranean collecting events together (47 field days) result in a much lower final percentage of new species (52.6%; table 1) that does also not seem to have reached stability (supplementary table S5).

Using the raw data, the global raw cumulative percentage of new species is 75.1% (517 of 688 species; table 1). The mean percentage of the five geographic regions is 71.2%. With a baseline of 1,001 species formally described as of the end of 2008 and ignoring undetected species, these percentages imply global estimates of about 3,500 and 4,000 species, respectively, of Pholcidae (table 2).

Accounting for undetected species in the five megatransects results in slightly higher percentages when using date by species incidence matrices (mean 76.0%), but

significantly higher percentages when using locality by species incidence matrices (mean 84.2%) (table 2). These percentages result in global species estimates of about 4,200 with a 95% confidence interval of (3,300, 5,000) and 6,300 species with a 95% confidence interval of (4,000, 8,600), respectively. The relatively high sampling variation associated with global species richness based on locality data is mainly due to the relatively sparse data in some localities in the subtropical and Mediterranean region; see supplementary table S9. The currently known 1,700 species of Pholcidae may thus represent no more than about 25–40% of the total species richness.

Discussion

The assessment of global species richness inevitably depends on the use of estimation and extrapolation based on different assumptions (Colwell & Coddington, 1994). Among the many methodological approaches devised over the last decades (Stork, 1997; Mora et al., 2011; Costello et al., 2012; Scheffers et al., 2012; Iknayan et al., 2014; Stork et al., 2015), diversity ratios are intuitively appealing in their

TABLE 2 Summary of the average percentages of new species and estimates of total species richness, using (a) raw numbers and (b) the *Chao2 estimates* (from table 1); s.e. = standard error; C. I. = 95% confidence interval

	% New Species	Estimate of Total Species Number
(a) Raw numbers		
Mean %	71.2%	~3,500
Cumulative %	75.1%	~4,000
(b) <i>Chao2 estimates</i>		
Date by Species	76.0% (s.e. 2.56%)	~4,200 (s.e. 443)
	C. I. (71.0%, 81.0%)	C. I. (3,300, 5,000)
Locality by Species	84.2% (s.e. 2.98%)	~6,300 (s.e. 1,189)*
	C. I. (78.3%, 90.0%)	C. I. (4,000, 8,600)*

* Large s.e. and wide C. I. due to sparse data in the subtropical and Mediterranean region; see supplementary table S9.

simplicity (Hodkinson & Casson, 1991; Grassle & Maciolek, 1992) but have been heavily criticised. Depending on the specific method, their main shortcomings are the unjustified extrapolation from regional to global patterns and the assumption of homogeneous taxonomic knowledge (Colwell & Coddington, 1994; Stork, 1997; Briggs, 1999; Lambshhead & Boucher, 2003; Scheffers et al., 2012; Costello, 2016). We reduce the problem of extrapolating from regional to global species richness by using field data from several continents. There are still necessarily major sampling gaps (e.g., the Andes, Australia, mainland Asia), but we think that our sample is not severely biased for two reasons. First, table 1 shows that for locality by species matrices, the estimated percentages of new species across the five regions are remarkably similar. Second, except for the subtropical and Mediterranean regions, the two types of data (date by species and locality by species) in each of the other four (tropical) regions also give generally comparable estimated percentages of new species. The pronounced difference in the percentages between the two types of data in the subtropical and Mediterranean regions is mainly due to the relatively sparse data in some localities in that region (supplementary tables S5 and S9).

We are inclined to recommend in our particular case the use of the results based on locality by species matrices. From a theoretical perspective, the *Chao2* estimator is nearly unbiased if *rare* species (specifically, undetected species and unique species in samples) have roughly the same detection probabilities in each sampling unit (Chao & Colwell, 2017). This condition may be more likely fulfilled in a locality than in a field day. This is because the search in each locality continued until the subjective impression was formed that all species had been detected, regardless of the amount of time taken to reach that point.

Although we are confident that our estimate is in the right order of magnitude, we acknowledge a series of potential biases that may affect our estimates and that need further study. First, percentages of undescribed species in the Pholcidae samples were higher in tropical than in other regions. Most collecting events were in tropical regions, possibly biasing the final global value. The impact of this bias on the global estimate is probably low for two reasons: (1) Most species of Pholcidae are tropical (fig. 3). Thus, the low percentage of new species in most non-tropical samples has little impact on the global value. (2) The Mediterranean and subtropical localities sampled herein are biased towards well-sampled areas. The USA, the Balkans, Crete, Turkey, and Israel have been sampled for many decades by many people and percentages of new species are thus low as expected. By contrast, Oman is less well studied and the percentage of new species was 100% (9 of 9). A more unbiased sample of non-tropical localities would probably bring the percentage of new species closer to that found in tropical regions.

Second, a bias may result if unknown or missing species are functionally different from known and described species (Scheffers et al., 2012; Costello et al., 2015; Stork et al., 2015). Our samples are probably biased towards large-size widespread species and against small litter-dwelling endemics. In a previous analysis of the Atlantic Forest (Brazil) samples, 'new' species had significantly smaller ranges than 'old' species (Huber, 2014), and a bias towards large widespread species and against small-scale endemics is a default and occurs even in well-studied faunas (Scheffers et al., 2012; Essl et al., 2013). However, 'old' species in the Pholcidae samples were mostly or entirely collected by generalist collectors, while all 'new' species were collected during campaigns focused on Pholcidae and their microhabitats,

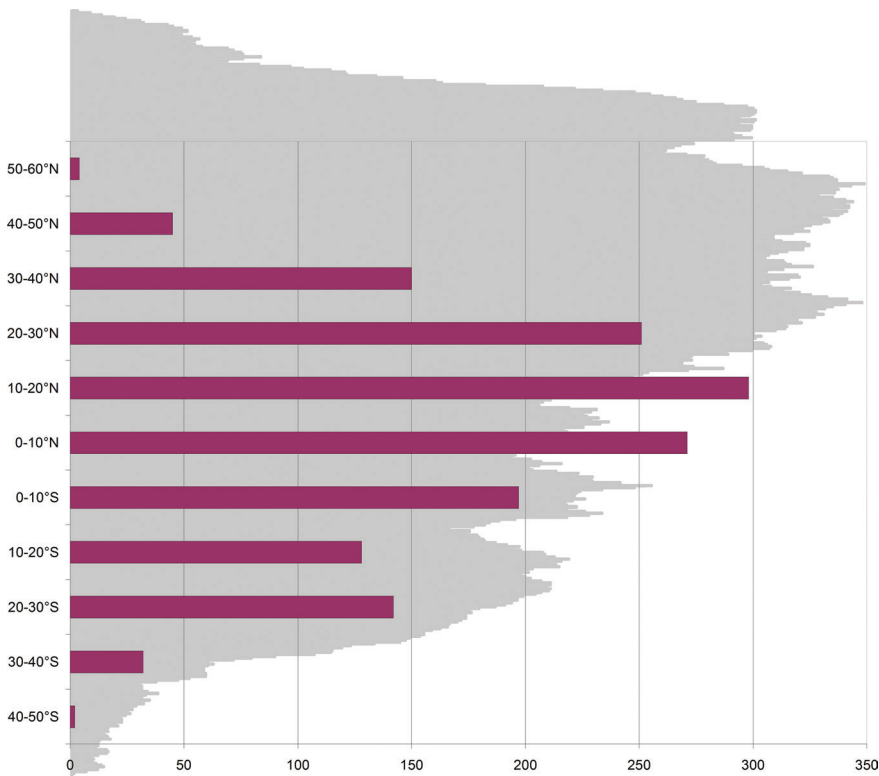


FIGURE 3 Latitudinal distribution of currently valid species of Pholcidae. Numbers of Pholcidae species (x-axis) known from different latitudes (y-axis; N, north; S, south), with the land mass distribution shown in grey (from www.ecoclimax.com; excluding Antarctica). Pholcidae species richness is slightly shifted towards the north, possibly as a result of the unbalanced land masses and/or taxonomists' biases, but most diversity is in tropical regions.

with special attention to the leaf litter, suggesting that this bias is probably not severe. In addition, the percentage of singletons (species represented by one specimen only) found in our samples was much lower than in large tropical arthropod surveys. For example, only 4.5% of the 132 species collected in the Brazilian Atlantic Forest were represented by singletons, while the average in large tropical arthropod surveys is at 32% (Coddington et al., 2009).

Third, the negative impact of synonyms and misidentifications on species richness estimates is widely acknowledged but the magnitude of the problem is far from clear (Solow

et al., 1995; Alroy, 2002; Scheffers et al., 2012). The problem is potentially severe in Big Data analyses that rely on digitized museum data (rather than on taxonomic revisions) without much possibility of verifying the quality of the data (Meier & Dikow, 2004; Wheeler, 2004). In our small sample of only 688 species of a single spider family, all of which were checked by a taxon expert for this particular group, synonyms and misidentifications are very unlikely to have a measureable impact on the general conclusion.

Further biases are likely to affect the estimate but are difficult to evaluate. Cryptic species (i.e., species that are morphologically

indistinguishable) might drastically elevate species richness estimates (e.g., Mutanen et al., 2013; Brehm et al., 2016; Larsen et al., 2017), but preliminary data on pholcid spiders suggest that cryptic species in this group are the exception rather than the rule (Astrin et al., 2006; Huber et al., 2010; Huber & Dimitrov, 2014). The use of different species concepts is likely to result in different numbers of species (Wheeler & Meier, 2000; Barrowclough et al., 2016). However, there is no reason to expect that the ratio of known to unknown species will be affected. Thus, different species concepts will change the numbers but not discredit the method.

Acknowledgments

We thank R.K. Colwell, J. Miller, and an anonymous referee for helpful comments on a previous version of the manuscript. The German Research Foundation is acknowledged for providing financial support to BAH (DFG Project HU 980/11-1). AC is supported by Taiwan Ministry of Science and Technology under Contract 106-2628-M007-01.

Supplementary material

Supplementary material is available online at: <https://brill.figshare.com/s/1e21b0c0784e2aac9121>

References

- Agnarsson, I., Coddington, J.A. & Kuntner, M. (2013) Systematics – progress in the study of spider diversity and evolution. In: D. Penney (Ed.) *Spider Research in the 21st Century*, pp. 58–111. Siri Scientific Press.
- Alroy, J. (2002) How many named species are valid? *PNAS* 99, 3706–3711. doi:10.1073/pnas.062691099.
- Astrin, J.J., Huber, B.A., Misof, B. & Kluetsch, C.F.C. (2006) Molecular taxonomy in pholcid spiders (Pholcidae, Araneae): evaluations of species identification methods using CO1 and 16S rRNA. *Zool. Scr.*, 35, 441–457. doi:10.1111/j.1463-6409.2006.00239.x.
- Barrowclough, G.F., Cracraft, J., Klicka, J. & Zink, R.M. (2016) How many kinds of birds are there and why does it matter? *PLOS ONE*, 11, e0166307. doi:10.1371/journal.pone.0166307.
- Brehm, G., Hebert, P.D.N., Colwell, R.K., Adams, M.-O., Bodner, F., Friedemann, K., Möckel, L. & Fiedler, K. (2016) Turning up the heat on a hotspot: DNA barcodes reveal 80% more species of geometrid moths along an Andean elevational gradient. *PLOS ONE*, 11, e0150327. doi:10.1371/journal.pone.0150327.
- Briggs, J.C. (1999) Marine species diversity. *Bioscience*, 49, 351. doi:10.2307/1313626.
- Caley, M.J., Fisher, R. & Mengersen, K. (2014) Global species richness estimates have not converged. *TREE*, 29, 187–188. doi:10.1016/j.tree.2014.02.002.
- Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, 43, 783–791. doi:10.2307/2531532.
- Chao, A. & Chiu, C.H. (2016) Nonparametric estimation and comparison of species richness. In: *eLS*. Wiley Online Reference in the Life Sciences. John Wiley & Sons, Ltd, Chichester. doi:10.1002/9780470015902.a0026329.
- Chao, A. & Colwell, R.K. (2017) Thirty years of progeny from Chao's inequality: estimating and comparing richness with incidence data and incomplete sampling. *SORT-Stat. Oper. Res. T.*, 41, 3–54.
- Chao, A., Gotelli, N., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K. & Ellison, A.M. (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.*, 84, 45–67. doi:10.1890/13-0133.1.
- Coddington, J.A., Agnarsson, I., Miller, J.A., Kuntner, M. & Hormiga, G. (2009) Undersampling bias: the null hypothesis for singleton species

- in tropical arthropod surveys. *J. Anim. Ecol.*, 78, 573–584. doi:10.1111/j.1365-2656.2009.01525.x.
- Coddington, J.A. & Levi, H.W. (1991) Systematics and evolution of spiders (Araneae). *Annu. Rev. Ecol. Syst.*, 22, 565–592. doi:10.1146/annurev.es.22.110191.003025.
- Colwell, R.K. & Coddington, J.A. (1994) Estimating terrestrial biodiversity through extrapolation. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 345, 101–118. doi:10.1098/rstb.1994.0091.
- Colwell, R.K., Mao, C.X. & Chang, J. (2004) Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology*, 85, 2717–2727. doi:10.1890/03-0557.
- Colwell, R.K., Chao, A., Gotelli, N.J., Lin, S.Y., Mao, C.X., Chazdon, R.L. & Longino, J.T. (2012) Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *J. Plant Ecol.*, 5, 3–21. doi:10.1093/jpe/rtr044.
- Costello, M.J. (2016) Parasite rates of discovery, global species richness and host specificity. *Integr. Comp. Biol.*, 56, 588–599. doi:10.1093/icb/icw084.
- Costello, M.J. & Chaudhary, C. (2017). Marine biodiversity, biogeography, deep-sea gradients, and conservation. *Curr. Biol.*, 27, R511–R527. doi:10.1016/j.cub.2017.04.060.
- Costello, M.J., Lane, M., Wilson, S. & Houlding, B. (2015) Factors influencing when species are first named and estimating global species richness. *Global Ecology and Conservation*, 4, 243–254. doi:10.1016/j.gecco.2015.07.001.
- 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. doi:10.1093/sysbio/syro80.
- Costello, M.J., Wilson, S. & Houlding, B. (2013) More taxonomists describing significantly fewer species per unit effort may indicate that most species have been described. *Syst. Biol.*, 62, 616–624. doi:10.1093/sysbio/syto24.
- Eberle, J., Dimitrov, D., Valdez-Mondragón, A. & Huber, B.A. (2018) Microhabitat change drives diversification in pholcid spiders. *BMC Evol. Biol.*, 18, 141. doi:10.1186/s12862-018-1244-8.
- Essl, F., Rabitsch, W., Dullinger, S., Moser, D. & Milasowszky, N. (2013) How well do we know species richness in a well-known continent? Temporal patterns of endemic and widespread species descriptions in the European fauna. *Global Ecol. Biogeogr.*, 22, 29–39. doi:10.1111/j.1466-8238.2012.00787.x.
- Grassle, J.F. & Maciolek, N.L. (1992) Deep-sea richness: regional and local diversity estimates from quantitative bottom samples. *Am. Nat.*, 139, 313–341. doi:10.1086/285329.
- Hamilton, A.J., Basset, Y., Benke, K.K., Grimbacher, P.S., Miller, S.E., Novotný, V., Samuelson, G.A., Stork, N.E., Weiblen, G.D. & Yen, J.D. (2010) Quantifying uncertainty in estimation of tropical arthropod species richness. *Am. Nat.*, 176, 90–95. doi:10.1086/652998.
- Hamilton, A.J., Basset, Y., Benke, K.K., Grimbacher, P.S., Miller, S.E., Novotný, V., Samuelson, G.A., Stork, N.E., Weiblen, G.D. & Yen, J.D. (2011) Correction. *Am. Nat.*, 177, 544–545. doi:10.1086/659643.
- Hodkinson, I.D. & Casson, D. (1991) A lesser predilection for bugs: Hemiptera (Insecta) diversity in tropical rain forests. *Biol. J. Linn. Soc. Lond.*, 43, 101–109. doi:10.1111/j.1095-8312.1991.tb00587.x.
- Huber, B.A. (2011) Phylogeny and classification of Pholcidae (Araneae): an update. *J. Arachnol.*, 39, 211–222. doi:10.1636/CA10-57.1.
- Huber, B.A. (2014) Progress and prospects in taxonomy: what is our goal and are we ever going to reach it? *J. Arachnol.*, 42, 142–147. doi:10.1636/CK13-69.1.
- Huber, B.A. (2015) Small scale endemism in Brazil's Atlantic Forest: 14 new species of *Mesabolivar* (Araneae, Pholcidae), each known from a single locality. *Zootaxa*, 3942, 1–60. doi:10.11646/zootaxa.3942.1.1.
- Huber, B.A. (2018) Pholcidae – the longest legs in the web. Online: www.pholcidae.de. Accessed 1 September 2018.

- Huber, B.A. & Carvalho, L.S. (2019) Filling the gaps: descriptions of unnamed species included in the latest molecular phylogeny of Pholcidae (Araneae). *Zootaxa*, 4546, 1–96. doi:10.11646/zootaxa.4546.1.1.
- Huber, B.A. & Dimitrov, D. (2014) Slow genital and genetic but rapid non-genital and ecological differentiation in a pair of spider species (Araneae, Pholcidae). *Zool. Anz.*, 253, 394–403. doi:10.1016/j.jcz.2014.04.001.
- Huber, B.A., Fischer, N. & Astrin, J.J. (2010) High level of endemism in Haiti's last remaining forests: a revision of *Modisimus* (Araneae: Pholcidae) on Hispaniola, using morphology and molecules. *Zool. J. Linn. Soc.*, 158, 244–299. doi:10.1111/j.1096-3642.2009.00559.x.
- Huber, B.A., Le Gall, P. & Mavoungou, J.F. (2014) Pholcid spiders from the Lower Guinean region of Central Africa: an overview, with descriptions of seven new species (Araneae, Pholcidae). *Eur. J. Taxon.*, 81, 1–46. doi:10.5852/ejt.2014.81.
- Iknayan, K.J., Tingley, M.W., Furnas, B.J. & Beissinger, S.R. (2014) Detecting diversity: emerging methods to estimate species diversity. *TREE*, 29, 97–106. doi:10.1016/j.tree.2013.10.012.
- Lamshead, P.J.D. & Boucher, G. (2003) Marine nematode deep-sea diversity – hyperdiverse or hype? *J. Biogeogr.*, 30, 475–485. doi:10.1046/j.1365-2699.2003.00843.x.
- 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. doi:10.1086/693564.
- Meier, R. & Dikow, T. (2004) Significance of specimen databases from taxonomic revisions for estimating and mapping the global species diversity of invertebrates and repatriating reliable specimen data. *Conserv. Biol.*, 18, 478–488. doi:10.1111/j.1523-1739.2004.00233.x.
- 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, 1–8. doi:10.1371/journal.pbio.1001127.
- Mutanen, M., Kaila, L. & Tabell, J. (2013) Wide-ranging barcoding aids discovery of one-third increase of species richness in presumably well-investigated moths. *Sci. Rep.*, 3, 2901. doi:10.1038/srep02901.
- Platnick, N.I. (1999). Dimensions of biodiversity: targeting megadiverse groups. In: J. Cracraft & F.T. Grifo (Eds) *The Living Planet in Crisis: Biodiversity Science and Policy*, pp. 33–52. Columbia University Press, New York.
- Platnick, N.I. & Raven, R.J. (2013) Spider systematics: past and future. *Zootaxa*, 3683, 595–600.
- 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. *TREE*, 27, 501–510. doi:10.1016/j.tree.2012.05.008.
- Solow, A.R., Mound, L.A. & Gaston, K.J. (1995). Estimating the rate of synonymy. *Syst. Biol.*, 44, 93–96. doi:10.2307/2413485.
- Stork, N.E. (1997) Measuring global biodiversity and its decline. In: M.L Reaka-Kudla, D.E. Wilson & E.O. Wilson (Eds) *Biodiversity II: Understanding and Protecting our Biological Resources*, pp. 41–68. J. Henry Press, Washington, DC.
- Stork, N.E., McBroom, J., Gely, C. & Hamilton, A.J. (2015) New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. *PNAS*, 112: 7519–7523. doi:10.1073/pnas.1502408112.
- Wheeler, Q.D. (2004). Taxonomic triage and the poverty of phylogeny. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 359, 571–583. doi:10.1098/rstb.2003.1452.
- Wheeler, Q.D. & Meier, R. (2000) *Species Concepts and Phylogenetic Theory: a Debate*. Columbia University Press, New York.
- World Spider Catalog (2019) World Spider Catalog. Version 19.5. Natural History Museum Bern, Online: <http://wsc.nmbe.ch>. Accessed 10 January 2019. doi:10.24436/2.