Comprehensive inventory of true flies (Diptera) at a tropical site

Brian V. Brown et al.#

Estimations of tropical insect diversity generally suffer from lack of known groups or faunas against which extrapolations can be made, and have seriously underestimated the diversity of some taxa. Here we report the intensive inventory of a four-hectare tropical cloud forest in Costa Rica for one year, which yielded 4332 species of Diptera, providing the first verifiable basis for diversity of a major group of insects at a single site in the tropics. In total 73 families were present, all of which were studied to the species level, providing potentially complete coverage of all families of the order likely to be present at the site. Even so, extrapolations based on our data indicate that with further sampling, the actual total for the site could be closer to 8000 species. Efforts to completely sample a site, although resource-intensive and time-consuming, are needed to better ground estimations of world biodiversity based on limited sampling.

Correspondence and requests for materials should be addressed to B.V.B. (email: bbrown@nhm.org)

#A full list of authors and their affiliations appears at the end of the paper.
Assessing the richness of the world’s biodiversity has long been a goal of biologists. For many groups of organisms, this richness, whether measured in species, interactions, niches, biomass, morphologies, natural histories, behaviors, or other scientific currency, hinges on the megadiverse tropics, especially the New World tropics, where for many groups the greatest number of the world’s species occurs. May provocatively stated that what was needed was “to assemble a team of taxonomists, with a comprehensive range of expertise, and then make a rough list of all the species found in one representative hectare in the tropical rain forest... Until this is done, I will not trust any estimate of the global total of species.” Indeed, estimated numbers of world species have continued to range from stunningly large totals, such as 30–100 million species to modest numbers of 5–10 million species.

Insects make up the bulk of the world’s known species, outside of prokaryotes and fungi, in which species-level assignments are not clear in many taxa. Projections about terrestrial diversity largely hinge on the status of four major groups: Coleoptera (beetles), Diptera (true flies), Hymenoptera (wasps, bees, ants), and Lepidoptera (moths and butterflies). Coleoptera comprise the largest number of named species, but Diptera are possibly the least-studied of these megadiverse groups. Each of the four groups is far more diverse in comparison to more popular and intensively studied vertebrates. For example, the fly family Tipulidae (crane flies) alone includes about the same number of named species as all mammals and birds combined.

For entomologists, the goal of understanding more than a small fragment of tropical diversity has heretofore been largely unattainable, due to the shortage of taxonomists, the dearth of funding, challenging curatorial techniques (such as slide mounting), and the overwhelming numbers of unknown species. Relative to the needs of science, collections are small and limited in scope such that Bickel found collecting intensity to be more correlated with species richness than any inherent characteristics of collecting sites. The number of well-known mainland tropical areas is virtually none. Instead, numbers of species in tropical inventories are estimated by taking small samples of faunas (often including just a few families), and extrapolating against some putatively known larger group or area. Unfortunately, with such extrapolations, there is little anchoring in verifiable data. Even more poorly known is the range of morphologies, life histories, hosts, parasitoids, and other data that make an inventory more than just a number.

One of the world’s best-known biotas is that of Great Britain, where generations of naturalists have searched the landscape looking for, recording, and classifying their local biodiversity. There, ~24,000 species of insects are found, of which about 7,000 species are Diptera, many more than 4000 Coleoptera. In comparison, for all the research that has been done in the mainland tropics (some islands have been more extensively surveyed), almost no single site has been successfully surveyed by a group of scientists for a major group of biodiversity-rich insects. A single exception is a beetle inventory of Dumoga-Bone National Park in Sulawesi, Indonesia, for which a large effort was made to obtain all species.

In the face of the challenge of understanding tropical diversity of Diptera, a group of dipterists decided to determine all species of Diptera collected from a single site, a cloud forest in Zürquí de Moravia, Costa Rica (hereafter referred to as Zürquí). To make this project feasible, the time scale for collecting (one year), geographical scale (four hectares), and collecting protocols were strictly constrained.

The ZADBI Zurquí All Diptera Biodiversity Inventory project (http://www.tropicalflies.net/) is the first successful modern effort to directly measure species richness of any megadiverse order of insects in a mainland tropical area. We found 4332 fly species based on direct evidence, which is important for other inferences on conservation of natural tropical and non-tropical environments in the world, adjustments of biodiversity study protocols, and understanding the funding needs for study and estimation of the group worldwide.

Results

Number of specimens. The ZADBI collecting effort resulted in 613 sampling events at Zürquí. From the many hundreds of thousands of Diptera collected, 52,947 were available for identification; of these 41,001 were identified. Many thousands of others were examined in alcohol but not identified in the database because they were perceived to be duplicates, or were the wrong sex for identification.

Number of species. From the 41,001 specimens, 4,332 species of Diptera were identified (Supplementary Data 1). Seventy-three of the world’s approximately 160 families were collected. We suspect that a small number of other families could theoretically be present (e.g. Platypeza sp., family Platypezidae were collected at Zürquí in 1995), but in general, most families expected to be present were found and treated.

Largest families. The most species-rich family at this site was the Cecidomyiidae (gall midges), with 800 species recorded (Table 1). As one of the least-studied families of flies in the world (relative to the number of species present), Cecidomyiidae also contain some of the smallest and most fragile specimens. Great care and special handling were necessary to collect, separate and slide mount these specimens without causing damage that would prevent their identification. Only a small fraction of the enormous amount of material collected by the various traps could be mounted and identified, and further work would undoubtedly yield even more species, perhaps allowing the total to exceed 1000. Cecidomyiidae generally have not been included in inventories that use trapping because previous research on plant-associated species of this family concentrated on obtaining all life stages through rearing. The ZADBI project is nearly unique in this regard, as we found that specimens of Cecidomyiidae have abundant, previously unused, structural characters, especially in the male genitalia, that make their identification possible. We object to the idea that alone among the Diptera, this family cannot be identified using morphology; the feasibility of this approach is demonstrated by the 800 species recognized herein.

| Table 1 Number of species from fifteen most species-rich Diptera families at Zürquí (from the current study) |
| Family | Number of species |
|--------|------------------|
| Cecidomyiidae | 800 |
| Phorididae | 404 |
| Tachinidae | 286 |
| Mycetophilidae | 268 |
| Tipulidae | 225 |
| Drosophilidae | 219 |
| Sarcophagidae | 204 |
| Ceratopogonidae | 200 |
| Dolichopodidae | 178 |
| Psychodidae | 171 |
| Chironomidae | 138 |
| Muscidae | 120 |
| Agromyzidae | 117 |
| Lavauidae | 116 |
| Syrphidae | 93 |
A notable exception to the lack of cecidomyiids in inventories is Hebert et al.22, who identified taxa of Canadian insects collected by Malaise trapping using DNA barcodes. They found Cecidomyiidae to be significantly underestimated, and extrapolated a worldwide total of 1.8–2.0 million species in this family alone.

The second most species-rich family was Phoridae, with 407 species (Table 1). Like cecidomyiids, phorids were represented in Malaise trap samples in huge numbers, with 100,000 or more specimens from Malaise trap 1 alone. We were unable to process all of this material, but the 407 species recognized is probably well below the actual total present. The third to seventh largest families were Tachinidae, Mycetophilidae, Tipulidae, Drosophilidae, Sciaridae, and Ceratopogonidae, groups that were also documented as abundant and understudied by Brown23 in other Neotropical Malaise trap samples.

The Zurquí ranking of the top 15 most species-rich families (Table 1) differs from that of the best known fauna, that of Great Britain17, and to that of the current world list24. This suggests that the current understanding of species richness per family has important biases, and that especially the world list is dominated by the idiosyncratic way that taxonomy develops. Families with large, showy species; those of agricultural, medical, and veterinary importance; and in one case, those that were the subjects of study by a determined and almost superhumanly productive individual (C.P. Alexander describing over 10,000 Tipulidae), led to taxonomic favorites being ranked as most species-rich (Table 2).

Our results suggest that the world list in particular is highly deficient in species of Cecidomyiidae, Phoridae, and other taxa whose lines connecting names from Zurquí to the world fauna in Fig. 1 (on the right side) slope downward to the right.

**Table 2 Fifteen most species-rich Diptera families**

| Family            | Number of described species |
|-------------------|----------------------------|
| Tipulidae s.l.    | 15,457                     |
| Tachinidae        | 8500                       |
| Asilidae          | 7479                       |
| Dolichopodidae    | 7236                       |
| Chironomidae      | 7054                       |
| Ceratopogonidae   | 6267                       |
| Cecidomyiidae     | 6203                       |
| Syrphidae         | 6016                       |
| Muscidae          | 5210                       |
| Bombyliidae       | 4946                       |
| Tephritidae       | 4911                       |
| Tabanidae         | 4406                       |
| Drosophilidae     | 4315                       |
| Mycetophilidae    | 4164                       |
| Phoridae          | 4105                       |

Numbers are on a worldwide basis48, with updated Cecidomyiidae49, Ceratopogonidae50, Drosophilidae51, Tephritidae (compiled by A.L. Norrbom), Tachinidae52, and Tipulidae53.

**Genus-level diversity**. Generic concepts are human constructs, but are useful when they represent monophyletic assemblages of species. The overall fauna at Zurquí is not dominated by large radiations of species but instead by many genera with a few species in each. For instance, 49% of the genera were represented by one species, 65% by one or two species. The mean number of species per genus is 4.63. These genera are not restricted to Zurquí; instead, they have relatives in other parts of the world25,26, notably other sites in Central America.

In contrast to the overall fauna, two genera are extremely diverse at Zurquí (and elsewhere): *Megaselia* Rondani with 240 species (Phoridae) and *Mycetophila* Meigen with 136 species (Mycetophilidae). Both were considered “open-ended”, nearly impossible groups by Bickel27, who declared them too diverse and too difficult to collect in their entirety for us to ever fully know them.

The phorid genus *Megaselia* represents 60% of the phorid fauna at Zurquí (Supplementary Data 1), but with much of the material still not identified due to the overwhelming number of specimens. The taxonomic status of this genus is uncertain because of questions of its monophyly28–30. The genus is...
Fig. 2 Species accumulation curves for fully extracted and identified families. Estimators from EstimateS: 

- **S(est)** Expected number of species in t pooled samples, given the reference sample(s); ICE Incidence-based Coverage Estimator; ACE Abundance-based Coverage Estimator.

a Malaise trap 1, b Malaise trap 2
Table 3 Results of species accumulation analysis of 49 fully-extracted families of Diptera

|                       | Malaise trap #1 | Increase factor | Malaise trap #2 | Increase factor |
|-----------------------|-----------------|-----------------|-----------------|-----------------|
| Number of specimens   | 4752            | —               | 291             | —               |
| Observed number of species | 590             | —               | 116             | —               |
| ACE                   | 1067            | 1.8             | 257             | 2.2             |
| ICE                   | 1095            | 1.9             | 297             | 2.6             |
| Chao 1                | 1068            | 1.8             | 252             | 2.2             |
| Chao 2                | 1054            | 1.8             | 273             | 2.4             |
| Mean value            | 1071            | 1.8             | 270             | 2.3             |

See Supplementary Data 2 for list
Increase factor = number of species estimated divided by the observed number of species
Estimators from EstimateS.35 (est) Expected number of species in t pooled samples, given the reference sample(s), ICE Incidence-based Coverage Estimator, ACE Abundance-based Coverage Estimator

Fig. 3 Rarefaction of species accumulation curves for Malaise traps 1 and 2

Extrapolations. Based on the Chao 1 formula, our data (S0obs = 3,925; #singletons = 1,777; #doubletons = 682) give an estimated total of 6647 species, 1.7 times the observed total. These data exclude the phorids, which were processed differently and all of which are singletons. If phorids were included, and their observed number of species (407) scaled in the same way, the projected number of phorids would be 689 species, and the estimate for the total Diptera fauna would be 7336 for the site.

None of the estimators for the sample-based species accumulation curves for the combined 49 families (for which all specimens were extracted and identified) reached an asymptote, although they began to converge on values (Fig. 2). Both graphs show the number of species accumulated to be well below the expected total, with estimators predicting 1071 species for Malaise trap #1 (1.8 times the 590 species actually collected) and 270 species, for Malaise trap #2 (2.3 times the 116 species actually collected) (Table 3). Rarefaction of the data from the two traps indicate that they were sampling a similar-sized fauna, albeit with Malaise trap #2 collecting more slowly (Fig. 3).

Discussion

Based on the sample-based estimates from the catch of our two Malaise traps, we found that our collecting underestimates the fauna by at least 1.7 times. Thus, our total of 4332 species could easily be 7,364 species (4332 × 1.7), or more likely even higher, since some of our estimates included only 49 families and did not include, among others, the megadiverse Cecidomyiidae or phorids. The estimated total is similar to that for the individual-based Chao 1 estimations of 7336 species.

Reasonable and predictive extrapolations of global diversity will become available only when more tropical faunas are interpreted at the species level, including those that are small-bodied or otherwise difficult to curate. Until then, we continue to be highly skeptical of projections, especially those at the lower end of the published ranges. For instance, a study of 0.48 hectares of a lowland tropical forest in Panama31,32 collected 6144 species of arthropods extrapolated using various models to suggest that the total reserve of 6000 hectares supported 25,000 species. Although including 102 researchers to identify an ecologically broad spectrum of taxa, their taxon sampling was actually quite low for the Diptera, which included only four families (compared to the 73 families in the current study). Of these, Dolichopodididae had 132 species, Asilidae 21, Scatopsidae 16 and Stratiomyidae 24 (178, 20, 22, 36 at Zurqui). This led them to suggest the presence of either 1754 or 1429 species of Diptera, depending on global or local comparisons. Even considering their more limited sampling area and different habitat in the lowlands, this contrasts strongly with the ZADBI collection of 4332 species and a total estimation of about 7300 species. It is likely that Diptera were seriously underestimated due to their constricted taxon sampling, much more so than their statement “…if the global species richness of Nematoecera is grossly underestimated, then our estimate of total species richness may be more than twice as high as the one reported here…” indicated. It is consequently uncertain what impact this has on their estimation of 25,000 arthropod species present at their site, as well as their larger extrapolated conclusion that “The robust estimates of local arthropod diversity derived in our study thus support previous estimates of global species richness [of 6.1 million species].”

Clearly, the Diptera are much more species-rich than represented in some previous estimates. A comparison of Zurqui diversity34 family by family with described species from Central America, Colombia, the entire Neotropical Region and the world, showed that current numbers of named species of Diptera are extremely low. Hebert et al., extrapolated from their collecting that there might be 50,000 species of Canadian Diptera. Considering the Canadian fauna to be 1% of global biodiversity, their extrapolation would lead to 5 million Diptera species worldwide, a number that is not inconceivable considering the diversity found at Zurqui. In addition, although no other sites have been similarly inventoried, there are indications that Zurqui is not
alone in being rich in species. Studies on individual groups of Phoridae, for instance, have uncovered faunas much larger than that of Zurquí, e.g., 127 species of Apocephalus Coquillett and 39 species of Dohrniphora Dahl in the lowlands at La Selva, Costa Rica, versus 26 Apocephalus and 9 Dohrniphora at Zurquí.

The only somewhat comparable data from other insect orders is the butterfly survey in Indonesia. Given that this survey took place over a much larger area (500 ha), including elevations from 200 to 1140 m, it is difficult to relate their preliminary number of 5649 species to that documented herein for a site two orders of magnitude smaller.

The species accumulation curves and projections (Figs. 2 and 3) for the ZADBI collections indicate that the entire fauna still has been inadequately sampled. Therefore, even with its tremendous number of specimens collected and identified, the ZADBI project fell short of documenting all Diptera of this site. This does not represent a failure of the project, as the ZADBI project has generated a large amount of baseline data for future tropical inventories, but a reflection of the difficulty of doing a proper survey. Reasons for the shortfall might include undersampling of the forest canopy, where much of tropical forest diversity is generated. Therefore, even with its tremendous scale, intensive inventories at other sites are needed to determine the distribution of the spectacular Diptera diversity and to fully document as much of it as possible before these species disappear.

Our protocol and results suggest that conservative estimates of world Diptera biodiversity based on much less thorough sampling are probably too low, not only for their underestimation of megadiverse Cecidomyiidae, but also for the legions of other small-to-large, undescribed flies. Furthermore, a tremendous number of evolutionary lineages are involved in this diversity, with low numbers of species per genus being the norm, indicating that similar diversity will be found elsewhere. Many species are small in size, meaning that fine morphological examination will be necessary to interpret character evolution, even as molecular methods begin to become more common in performing quantitative assessments of biodiversity (e.g., DNA barcoding or high throughput sequencing after homogenizing samples with a blender). Many more large-scale, intensive inventories at other sites are needed to determine the distribution of this spectacular Diptera diversity and to fully document as much of it as possible before these species disappear.

Data availability. The data underlying the species accumulation analyses are given in Supplementary Materials.

Received: 30 December 2017 Accepted: 20 February 2018
Published online: 22 March 2018

References

1. Robbins, R. K. & Opler, P. A. in: Biodiversity II: Understanding and Protecting our Biological Resources (eds Reaka-Kudla, M. L., Wilson, D. E., Wilson, E. O.) 333–354 (John Henry Press, Washington, DC, 1997).
2. Gaston, K. J. & Hudson, E. Regional patterns of diversity and estimates of global insect species richness. Biodivers. Conserv. 3, 493–500 (1994).
3. Chazot, N. et al. in Biodiversity Conservation and Phylogenetic Systematics: Preserving our Evolutionary Heritage in an Extinction Crisis (eds Pellens, R. & Grandcolas, P.) (Springer International Publishing, Cham, 2016).
4. May, R. M. How many species are there on earth? Science 247, 1441–1449 (1988).
5. Erwin, T. L. How many species are there?: revisited. Conserv. Biol. 5, 330–333 (1991).
6. Caley, M. J., Fisher, R. & Mengersen, K. Global species richness estimates have not converged. Trends Ecol. & Evol. 29, 187–188 (2014).
7. Adams, M., Raadik, T. A., Burridge, C. P. & Georges, A. Global biodiversity assessment and hyper-cryptic species complexes: more than one species of elephant in the room? Syst. Biol. 63, 518–533 (2014).
8. Stork, N. E. Insect diversity: facts, fiction and speculation. Biol. J. Linn. Soc. 35, 321–337 (1988).
9. Basset, Y. et al. Arthropod diversity in a tropical forest. Science 338, 1445–1448 (2012).
10. Basset, Y. et al. Arthropod diversity in a tropical forest. Science 338, 1445–1448 (2012).
11. Erwin, K. J. The magnitude of global insect species richness. Conserv. Biol. 5, 283–296 (1991).
12. O’degaard, F. How many species of arthropods? Erwin’s estimate revisited. Biol. J. Linn. Soc. 71, 583–597 (2000).
13. Stork, N. E., McRoon, J., Gely, C. & Hamilton, A. J. New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. Proc. Natl Acad. Sci. USA 112, 7519–7523 (2015).
14. Stork, N. E. How many species of insects and other terrestrial arthropods are there on Earth? Annu. Rev. Entomol. 63, 31–45 (2018).

Methods

Sampling protocol. Methods of the ZADBI project were previously outlined in detail, including rationale for selecting this site. In short, Zurquí was selected to maximize biodiversity, based on anecdotal data from collections made at the site over the 20 years since its discovery, and to minimize difficulties, being on private land and located a short distance from the collaborating Instituto Nacional de Biodiversidad (INBIO). A variety of sampling methods were used at a partially permanent stream. Three days per month, Malaise trapping was supplemented by light-interception trapping, light trapping, pan trapping, baiting, and emergence trapping. Three additional Malaise traps were operated to provide a specimen pool for families represented by low numbers. All Diptera were separated, sorted to family, then either curated, or for perceived duplicates of previously extracted species, left in alcohol. Curated specimens were fully prepared to our specifications, generally as dried or slide-mounted specimens, entered into our database, and distributed to each of us for detailed study. Species identifications were based on morphological criteria only. Once identified, names of species, or numbered morphotypes were added to our database (Supplementary Data 1).

Although not fully indicative of whether or not a species was previously described, we note that 93% of all species identifications were morphospecies, i.e., not immediately recognized by experts as named taxa. Further taxonomic work, in many cases full revisions, would be necessary to recognize how many undescribed species were present in the material.

Phoridae were treated slightly differently, in that samples were examined and only presumed novel species were removed, recorded, and entered into our database.

Species number estimates. An estimate of the number of species actually occurring at Zurquí was made using the Chao1 formula, where singletons is the number of species known from only one specimen and doubletons those only known from two specimens (this formula is only valid if doubletons > 0). All families and all samples were used for this estimation, except phorids, whose sorting only recorded singletons. To study the effectiveness of the two primary Malaise trap samples, trap-based species accumulation curves were developed with Estimates, using abundance data for 49 families of Diptera (less diverse families that could be fully extracted and identified: Agromyzidae, Anisopodidae, Anthomyiidae, Anthomyi- zidae, Asilidae, Asteidae, Aulacigasteridae, Bombyliidae, Calliphoridae, Chamaemyiidae, Clusio- ridae, Conopidae, Ctenostylinidae, Culicidae, Diastatidae, Dixidae, Ephyrididae, Fanniidae, Heleomyzidae, Inionyidae, Lonchaeidae, Lygisti- rhinidae, Micropezidae, Milichiidae, Neridiidae, Oedematiidae, Pericelliidae, Piophilidae, Pipunculidae, Pseudopomyzidae, Pyrgotidae, Rhagionidae, Rheophoridae, Richardiidae, Sarcophagidae, Scathophagidae, Scatopsidae, Sepsidae, Simuliidae, Stratiomyiidae, Syrphidae, Tabanidae, Tachinidae, Tanyopidae, Teprithridae, Uliidae, Xylomyidae, Xylophagidae) from 50 week-long Malaise trap samples from Malaise traps 1 and 2. The program considers each sample to be a unit for constructing a species accumulation curve, randomizes the samples and smooths the curves based on sampling within all samples (Supplementary Data 2).

Rarefaction of the two sample sets was performed using Next Online, which procedure compares samples for distinct sites and equalizes the sampling effort based on the number of specimens, rather than samples (Supplementary Data 3). The point at which the samples are joined based on the limitation of the number of species in the smaller sample, is shown by a large dot in Fig. 3.

Family classification follows that of Brown et al., other than the Empididae sensu lato, which are here treated as three families, namely Empididae, Brachystomatidae, and Hybotidae.
