CHAPTER TWELVE

# BIODIVERSITY RESEARCH BASED ON TAXONOMIC REVISIONS – A TALE OF UNREALIZED OPPORTUNITIES

Torsten Dikow[1], Rudolf Meier[2], Gaurav G. Vaidya[2] & Jason G. H. Londt[3]

[1]*Cornell University, Ithaca, USA and*
*American Museum of Natural History, New York, USA;*
[2]*National University of Singapore, Singapore;*
[3]*Natal Museum, Pietermaritzburg, South Africa and*
*University of KwaZulu-Natal, Pietermaritzburg, South Africa*

## Introduction

If we were to ask the average biologist or university administrator about his or her opinion on the relative importance of revisionary taxonomy and biodiversity research, the vast majority would consider taxonomy uninteresting or even unnecessary, while many would find biodiversity research interesting and important. After all, we live at a time when we do not know even within an order of magnitude how many species exist on our planet, where they are found, and whether they are threatened by extinction. Yet, here we will argue that these positions are incongruous, because without the help from revisionary taxonomy, biodiversity and conservation research will remain restricted to less than 10% of the known species diversity; i.e., mostly vascular plants, butterflies, mammals, and birds. Some will argue that the remaining 90% can be safely ignored because they are less glamorous and deserve less attention. Glamour may be important for conservation organizations when collecting donations from the public and arguing for the conservation of an area, but when it comes to scientific research in biodiversity and conservation biology it is important to also consider invertebrate species. It also appears from the literature, that biodiversity researchers are not avoiding invertebrate data because they are regarded as unimportant. Instead, such data are gener-

Diptera Diversity: Status, Challenges and Tools
(eds T. Pape, D. Bickel & R. Meier). © 2009 Koninklijke Brill NV.

ally regarded as unavailable. Here we will demonstrate how specimen lists from taxonomic revisions will provide this much needed access. We will furthermore argue that using these data will not only help biodiversity research. Analyzing specimen lists from revisions also creates new opportunities for taxonomists, who are now living in a scientific environment where they need to make their research more relevant to a larger audience and need to produce more publications with high immediate impact (Wheeler 2004).

The focus of this chapter is not yet another discussion of what biodiversity means, whether the inclusion of invertebrates is desirable, or how biodiversity can be preserved. Instead, we will use several Diptera examples for discussing how invertebrate data can be incorporated into quantitative biodiversity research. We will demonstrate how specimen data from taxonomic revisions can be used (1) to compare the species richness and levels of endemism of two or more areas, (2) to provide quantitative data for Red Lists of endangered species, and (3) to estimate the full species diversity in a clade. We will start by pointing out how much data are already available in the taxonomic literature and end with discussing the problems with using specimen data. Throughout the chapter we will use examples from the Asilidae (Diptera: Brachycera: Asiloidea). Robber flies are a diverse group of predatory insects (some 7,000 described species) that mostly catch other insects on the wing. The largest species diversity is found in arid and semiarid regions all over the world. In contrast to the many invertebrate groups that are only collected by a dedicated group of specialists, asilids are popular with amateur and professional entomologists alike. Many robber flies are large or conspicuous due to their habit of resting on exposed vegetation or on the ground. Furthermore, catching asilids poses a nice challenge gladly taken up by many collectors. Due to the combination of these factors, Asilidae collections are unusually large and collected by a diverse group of entomologists thus creating a more random specimen sample than is available for most other groups of invertebrates.

In discussing the use of specimen data from taxonomic revisions, we will rely on two data sets. One covers a large proportion of the sub-Saharan Asilidae. Londt (1977–2002, 37 publications) and Dikow (2000–2003, 3 publications) have revised and described 724 species of the approximately 1,500 described Afrotropical robber flies and we compiled 21,505 specimen records from these taxonomic revisions. The second data set was compiled for a revision of the Danish Asilidae fauna. This data set is unusual in that it contains 4,300 specimens for a relatively small fauna of only 30 species.

## 1. Taxonomic Revisions: How Much Information is Available?

The scientific literature contains a vast number of taxonomic revisions and most include specimen data. For example, in a recent search Meier & Dikow (2004) found that the *Zoological Record* listed more than 2,300 taxonomic revisions that were published between 1990–2002 and Gaston (1991) documented that more than 10,000 new species in the four hyper-diverse insect 'orders' were described between 1986–1989. In order to be able to produce distribution or range maps and to analyze the phenology or seasonality of species, taxonomists routinely collect data from the specimen labels. These labels often include, for example, locality, date/year of collection, collector, altitude, and ecological information. The data are generally available within the revision, sometimes in smaller font, as an appendix, or as an electronic supplement. We will here promote the use of these data and believe that they are preferable over data obtained by the currently more popular approach of digitizing label information from museum specimens:

1. The data from revisions are readily available and do not require specimen label digitization by non-specialists; they are thus more cost-effective and of higher quality because an expert can avoid many data transfer errors.

2. The specimens were identified by the best expert in the field; i.e., the taxonomist who is carrying out the taxonomic revision. He is often the only expert in the world who can correctly identify closely related species. Misidentifications which can be common in museum collections (Meier & Dikow 2004) are thus, as much as possible, avoided.

3. Specimens from many collections contribute label data to the specimen list in taxonomic revisions thus maximizing specimen coverage for a particular group.

4. Specimens in the 'unsorted' drawers are more likely to be included because specialists tend to sort through or borrow unsorted material from a large range of institutions.

5. Given the large number of published revisions, millions of label data are already available and there is no need to wait for museum digitization projects to be completed.

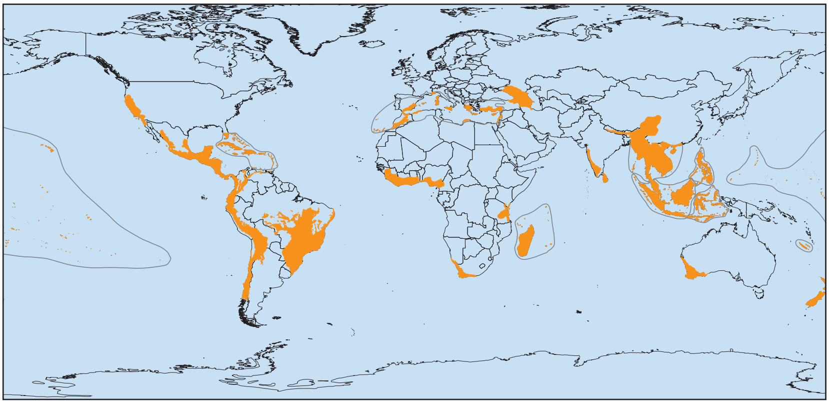

Figure 12.1. Map of the world with biodiversity hotspots in orange.

## 2. Use of Specimen Data for Comparing the Biodiversity of Conservation Areas

One of the main goals in conservation biology is to optimize the selection of conservation areas given that only very limited financial resources are available for protecting biodiversity. Special attention is usually given to areas with outstanding biodiversity whereby the latter is assessed using a variety of different criteria. In the older literature, raw species counts were often used, which had the undesirable effect that many species with small ranges were not covered by the selected regions. Today, more attention is paid to a variety of other criteria. For example, Myers *et al.* (2000) focused on maximizing the number of endemic species in habitats that have already lost most of their natural vegetation. In 1991, Humphries *et al.* argued for a combined approach including species richness, phylogenetic diversity, complementarity, and taxonomic distinctness, and Vane-Wright *et al.* (1991) proposed a taxic diversity measure based on complementarity analysis of faunas and floras. One year later, Platnick (1992) discussed the comparison of species richness and species composition (overlap) between three allopatric areas and favored an approach that would maximize the preservation of species assemblages. All these techniques have one element in common. They are critically dependent on having species distribution data. This also explains why invertebrates have been largely ignored in these analyses because it is commonly assumed that distribution data are not available.

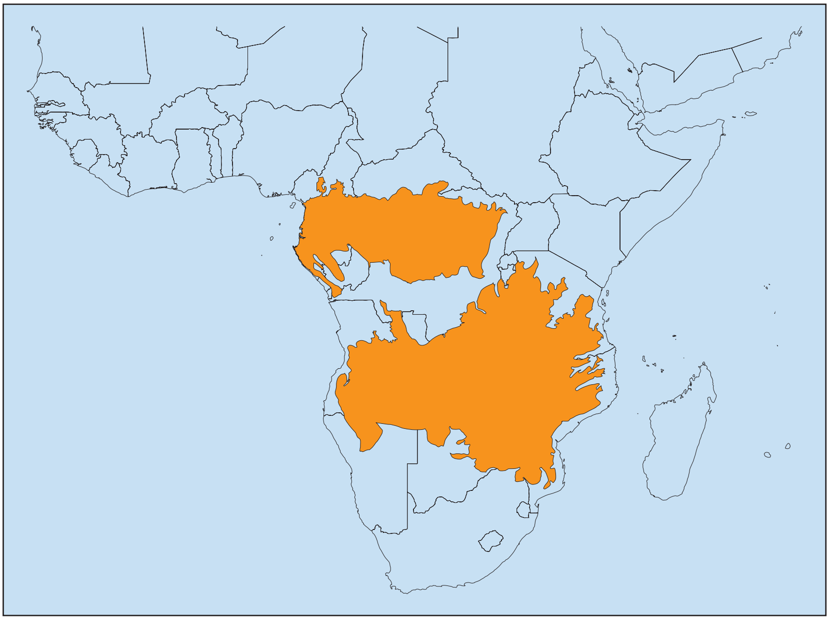

Figure 12.2. Map of sub-Saharan Africa with Wilderness Areas (Congo Basin and Miombo-Mopane) in orange.

### 2.1 *Assessing the validity of Myers* et al.*'s biodiversity hotspots for Diptera*

Many biologists have pondered the question where on our planet we find the highest and most threatened animal and plant diversity. Botanists were the first to provide answers because they have long accumulated detailed information on the distribution of vascular plant species published in numerous regional floras. Analyses of these data revealed global diversity centers (Barthlott *et al.* 1996) and biodiversity hotspots (Myers *et al.* 2000; Fig. 12.1). Barthlott *et al.* (1996) evaluated the number of species within a specific area of 10,000 km$^2$ and distinguished diversity regions by isotaxas; i.e. lines of equal species richness. Myers *et al.* (2000) went beyond just mapping species richness and also incorporated a conservation angle by considering species endemism and habitat loss. Myers *et al.* designated areas as biodiversity hotspots if they harbored at least 1500 endemic species of vascular plants (0.5% of the described species diversity on our planet) and if 70% of the habitat in the area had already been lost through human interference. Twenty-five biodiversity hotspots were originally defined and today these cover a combined area of only 1.4% of the Earth's land masses. Yet, they contain 44% of all flowering plants as endemic species. Even more remarkable was the finding that, although these hotspots were defined based on plant distributions, they were also hotspots for terrestrial tetrapods (Myers *et al.* 2000). When the distribution of mammals, birds, reptiles, and amphibians were mapped, 35% of all tetrapod species were endemic to the same biodiversity hotspots that had been defined based on plant distributions. The hotspots also performed well in protecting phylogenetic diversity. Some authors had argued that not all species are equal and that it is also important to protect phylogenetic diversity. Fortunately, the biodiversity hotspots were found to be home to a large proportion of the phylogenetic diversity in birds and primates (Sechrest *et al.* 2002).

In contrast to the biodiversity hotspots, some ecosystems of the world remain largely undisturbed and these were recently defined and termed 'wilderness areas' (Mittermeier *et al.* 2003; Fig. 12.2). They occupy 44% of all terrestrial habitats on Earth, have low human population densities (<5/$km^2$), and have lost less than 30% of their original vegetation. However, despite being large, Mittermeier *et al.* (2003) found that they harbor only 18% of all plant and 10% of all terrestrial vertebrate species as endemics; i.e., they are not very effective in providing a safe haven for a large proportion of vascular plant and terrestrial tetrapod species.

These studies have been very influential in defining conservation priorities and much funding is now channeled into the protection of biodiversity hotspots regardless of the fact that all this research only considers much less than 10% of the global species diversity. The obvious question is whether invertebrate diversity is also concentrated in these hotspots defined based on plants. In order to answer this question, we decided to test the biodiversity hotspots and wilderness areas in sub-Saharan Africa for robber flies based on all available specimen data from taxonomic revisions. We plotted the distribution of the 21,505 specimens representing 724 species of Afrotropical Asilidae on a map of sub-Saharan Africa (Fig. 12.3). Overall, 1,727 unambiguous localities were included (= geographic co-ordinates are known). As is evident from the map, the most comprehensive revisionary research has been conducted on the fauna of the Republic of South Africa, which was the focus of Jason Londt's studies. South Africa is also the home of two biodiversity hotspots *sensu* the original circumscription of the biodiversity hotspots by Myers *et al.* (2000; Fig. 12.3). These are the Succulent Karoo (SK) on the Atlantic west coast stretching north to southwestern Namibia and the Cape Floral Region (CFR)

 

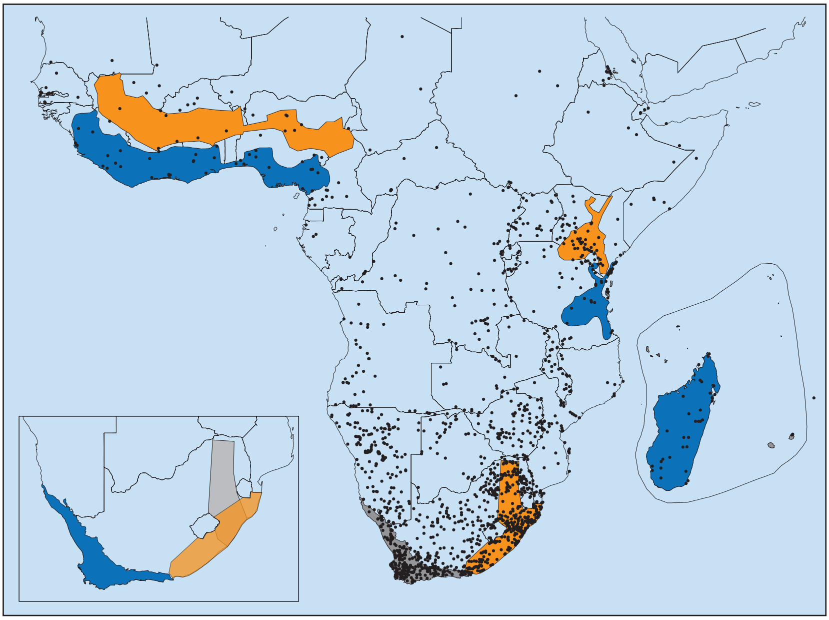

Figure 12.3. Map of sub-Saharan Africa with collecting localities (circles), Biodiversity hotspots in blue, and 'imaginary hotspots' in orange. Inset with detailed map of southern Africa with combined Cape Floral Region and Succulent Karoo biodiversity hotspot and two 'imaginary hotspots' ESA and SEA.

coinciding with the Mediterranean-type climate in southwestern South Africa. Other hotspots tested in our study are the Eastern Arc Mountains in Tanzania and Kenya and the Guinea Forests comprising the tropical rainforest belt along the Atlantic Ocean coast of western Africa (Fig. 12.3). Although Madagascar constitutes yet another sub-Saharan hotspot, it is here excluded because of its island status.

We find that of the 724 robber-fly species in this study, 295 (41%) are resident in at least one of the biodiversity hotspots, which combined occupy 8.5% of the surface area of sub-Saharan Africa. This number is unexpectedly high compared to the corresponding numbers for vascular plants (42.2%) and terrestrial vertebrates (26.4%). Note that the latter two numbers are actually overestimates because in the case of species overlap between several biodiversity hotspots, species are double-counted, which makes the performance of the hotspots for Asilidae even more impressive. However, when we only consider the endemic species among the hotspot residents, we find that only 149 (20.6%) fall into this category. This is a considerably lower proportion of endemics than the hotspot endemicity values for vascular plants (Africa: 42%, global: 44%) and terrestrial vertebrates (Africa: 29%, global: 35%).

At this stage one may be inclined to reject the biodiversity hotspots for Asilidae, but this first impression deceives. In evaluating an area, the absolute numbers may give the wrong impression and it is equally important to compare the performance of an area earmarked for conservation to other areas of the same size and shape that are not proposed for conservation. We thus created 'imaginary hotspots' that were chosen based on three criteria: (1) identical size to a real biodiversity hotspot, (2) similar geographic location, and (3) similar sampling intensity. We rejected random area selection because it would likely yield areas of incomparable sampling intensity. For the same reason, only 1–2 comparison areas were here identified for each hotspot. For the Eastern Arc and the Guinea Forests, we tested areas of identical size and shape situated north of the original hotspot (Fig. 12.3). For the southern African hotspots, we combined the adjacent hotspots Cape Floral Region (CFR) and Succulent Karoo (SK) and created two 'imaginary hotspots' in eastern South Africa (ESA — Eastern South Africa and SEA — Southeastern South Africa; Fig. 12.3). These were of identical size, positioned at similar latitude as the combined original hotspots, and included an area that had been particularly well sampled by the staff of the Natal Museum (Pietermaritzburg, South Africa). For each imaginary hotspot, we established the number of collecting events and counted the number of endemic and resident species and genera (subtracting species overlap; Table 12.1). For the ESA and SEA imaginary hotspots, we used the mean number of resident endemic species for comparison with the combined real hotspot. For the combined CFR and SK hotspots, we used the sum of endemic species for both hotspots, which underestimates the correct number because some species that are not endemic for either hotspot may be endemic for the combined area.

When using the imaginary hotspots as point of reference, we find that the real hotspots perform extremely well for Asilidae. The real hotspots house 57% more species than are present in the imaginary hotspots (284 vs. 181 species). More remarkably yet, the levels of endemism in the real hotspots are elevated by 424% (140 vs. 33 species) over the level in the comparison areas. Furthermore, the biodiversity hotspots contain three endemic genera while endemism at the generic level is absent in the imaginary hotspots (Table 12.1). In order to rule out that the performance differences are due to unequal sampling, we obtained unbiased values through data-set resampling. The imaginary hotspots had lower numbers of collecting events than the corresponding real hotspots (Eastern Arc, real: 76 vs. imaginary: 62; Guinea Forests, real: 42 vs. imaginary: 25; combined CFR and SK, real: 1,822 vs. imaginary ESA: 984 and SEA: 1,261). To correct for this bias we rarefied the larger data sets from the real hotspots ten times to match the size of the respective smaller data sets and again determined the average number of resident and endemic species based on the resampled data sets (Table 12.1). The number of resident species and endemic species still remained higher in the biodiversity hotspots in com-

Table 12.1. Comparison of resident and endemic species in biodiversity and 'imaginary' hotspots.

| | Biodiversity Hotspots | Biodiversity Hotspots | 'Imaginary' Hotspots | Resampled Imaginary Hotspots[3] |
|---|---|---|---|---|
| | Resident/ Endemic species | Resident/ Endemic genera | Resident/ Endemic species/ Resident genera | Resampled resident/ Endemic species |
| Madagascar | 12/9 | 9/2 | n.a. | n.a. |
| Eastern Arc | 26/3 | 11/0 | 31/5/17 | 23±1.7/2.9±0.3 |
| Guinea Forests | 16/3 | 7/0 | 11/1/5 | 10±1.4/2.0±0.8 |
| Succulent Karoo (SK) | 163/48 | 29/0 | n.a. | n.a. |
| Cape Floral Region CFR) | 168/51 | 29/1 | n.a. | n.a. |
| Combined CFR & SK | 248/134 | 33/3 | ESA[2]: 147/21/34 SEA[2]: 150/32/33 | ESA: 191±5/98±5 SEA: 212±7/113±5 |
| **Totals** | **295/114, 149[1]** | **53/5** | **n.a.** | **n.a.** |
| **Totals w/o Madagascar** | **284/105, 140[1]** | **53/3** | **181/33** | **230±12/ 110±8.8** |

[1] 114 = Sum of CFR & SK, 149 = Combined CFR & SK

[2] Two imaginary hotspots tested for the combined CFR & SK hotspot

[3] Hotspot dataset resampled to dataset size of imaginary hotspots

parison to the imaginary hotspots (resident: 230 in real hotspots vs. 181 in imaginary => +27%; endemic: 111 in real hotspot vs. 33 in imaginary hotspot => +336%; Table 12.1).

The performance difference between the real hotspots and the comparison areas is mainly due to the contribution of the two southern African hotspots (Cape Floral Region and Succulent Karoo) that contain several large radiations of plant lineages. One of the most speciose asilid genera world-wide, *Neolophonotus* Engel, which is confined to the Afrotropical Region, is especially species-rich in these two hotspots and of the 270 described species 70 are endemic in the hotspots.

The wilderness areas proposed by Mittermeier *et al.* (2003) are a different kind of conservation area that can be tested for invertebrates. These areas are generally very large and the two high-diversity wilderness areas in Africa (Congo Basin and Miombo-Mopane) comprise 11.9% of the surface area of sub-Saharan Africa (Fig. 12.2). On the other hand, they have only 7,900 (15.8%) endemic vascular plant species and 170 (3.9%) endemic terrestrial vertebrate species (Brooks *et al.* 2001). Our Asilidae data indicate a relatively high number of species that reside in the Congo Basin and the Miombo-Mopane woodlands (196 species, 27%), but only a few, 60 species (8%), are endemic to them. Although large in size, the wilderness areas are thus also relatively ineffective in protecting the sub-Saharan robber-fly species diversity. The small number of endemic species is especially surprising given that large areas have a higher chance of harboring a large number of endemics (Brooks *et al.* 2002).

Overall, we find that the biodiversity hotspots that were defined based on plant distribution data perform well not only for terrestrial tetrapods, but also for Asilidae. This correlation is all the more surprising because Asilidae are not phytophagous. For phytophagous insects, a correlation would intuitively have been expected, but robber-fly larvae as well as the imagines are predators with little evidence for prey specialization beyond size. It thus appears more likely that, for example, historical reasons (e.g., geology, fragmentation through climate change) account for the simultaneously high levels of endemism in the biodiversity hotspots in vascular plants, robber flies, and terrestrial vertebrates. Our example here demonstrates how data from taxonomic revisions can be used to test whether existing conservation priority areas have any relevance to invertebrates. Note that these tests were carried out based on published data that are freely available in the literature. Note also that the same set of techniques

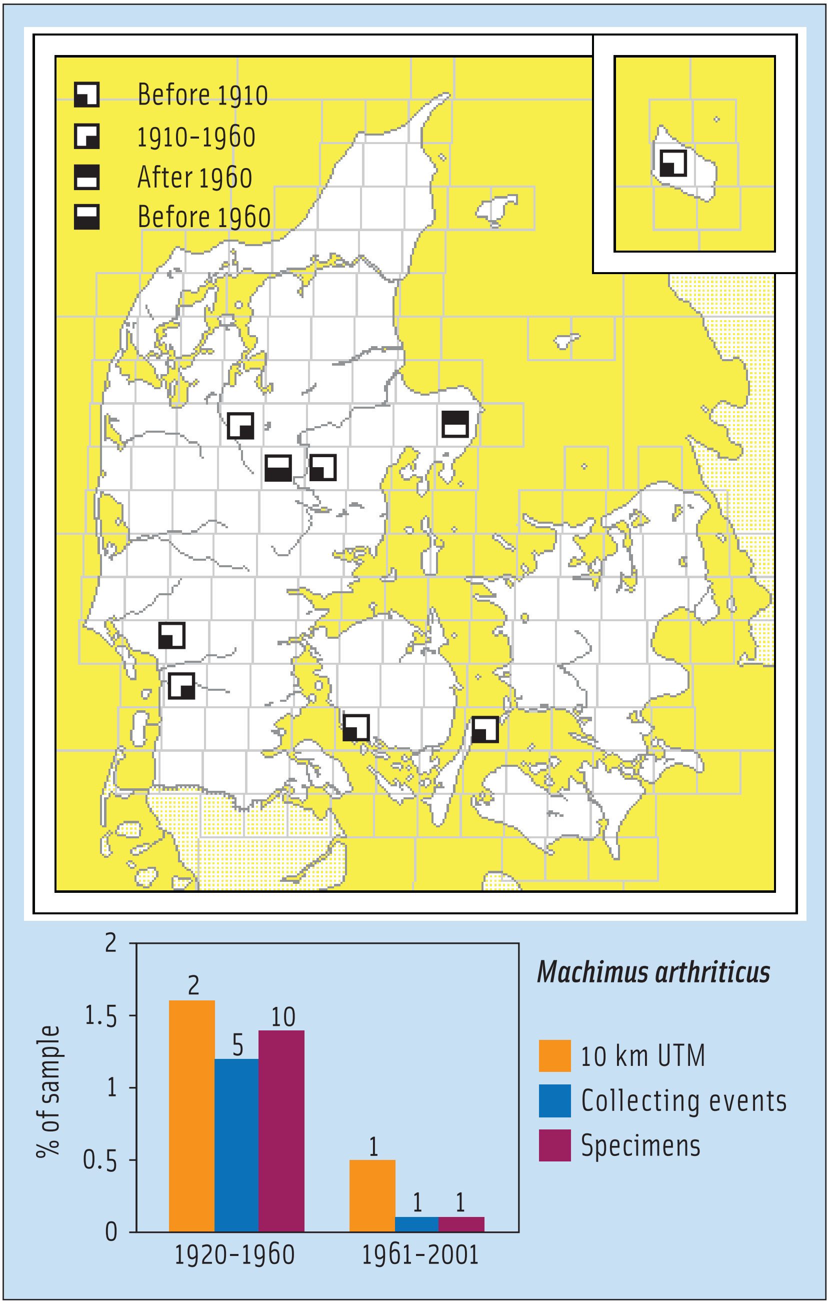


Figure 12.4. Collecting records for *Machimus arthriticus* (Zeller) in Denmark over two time periods (1920–1960; 1961–2001; inset: Bornholm).

can be used for assessing the validity of other conservation areas for invertebrates.

## 3. Use of Specimen Data for Proposing Red Lists

Red Lists for invertebrates are usually based on the specialists' guesses instead of quantitative data as demanded by the World Conservation Union (IUCN 2001), but see Red Lists for Finland and Sweden (Gärdenfors 2001, Gärdenfors *et al.* 2001). The main perceived problem is the lack of quantitative data for invertebrates, that supposedly prevented biologists from applying the IUCN criteria that require quantitative statements about species abundances, distributions, and/or probabilities of extinction. However, we will argue here that some taxonomic revisions contain enough information. Recently, Larsen & Meier (2004) revised the Danish Asilidae and proposed a Red List based on 4500 specimens for 30 species. The data set was divided into two time periods (1920–1960, 1961–2001) and for each the number of specimens, collecting events, and number of 10 $km^2$ UTM grids in which a particular species was found was determined. Changes in abundance were then evaluated after correcting for collecting effort by relating the records for a particular species to the overall collecting activity during the two time periods and the geographic regions that were sampled (for other methods, see Fagan & Kareiva 1997, Ponder *et al.* 2001).

After the correction, we still had to apply the IUCN criteria for regional Red Lists (Gärdenfors 2001, Gärdenfors *et al.* 2001). Of the criteria suggested by IUCN for ranking species, we believe only the 'geographic-range criterion' (criterion B, and rarely category D) can realistically be used for insects. It requires that it is documented that the 'extent of occurrence' or 'area of occupancy' (see IUCN 2001) is smaller than a defined size, whereby different threat categories have different size-thresholds. It is furthermore necessary to document that two of the following three phenomena apply to the species in question: (1) fragmented distribution or existence at few locations, (2) population decline (observed, inferred, or projected), and (3) extreme fluctuations in range or number of populations. The latter criterion probably can not be used for most insect groups, because there are not enough data to document such fluctuations. However, the first two criteria can be applied if enough data are available. We were able to rank all 30 Danish asilid species and found that seven species are now ‘critically endangered’ (CR), one species is ‘endangered’ (EN), four species are ‘vulnerable’ (VU), two species are ‘near threatened’ (NT), and 16 species are ‘least concern’ (LC). Below is a typical example for a species assessment:

*Machimus arthriticus* (Zeller): IUCN Criteria B1+2a,b: extent of occurrence and area of occupancy is probably less than 100 $km^2$ and 10 $km^2$ respectively (see Fig. 12.4); a: only one post-1950 locality is known; b: number of populations is in decline; evidence: the species has been collected five times in two 10 $km^2$ grids between 1920 and 1960. Afterwards, it has only been taken once and there is a decline in the number of UTMs, collecting events, and number of specimens. However, it must be pointed out that collecting in the grids with known occurrence has dramatically decreased after 1960 (1920–1960: 80 events, post-1960: 12). One might thus be inclined to consider the decline a sampling artifact, but between 1900 and 1920 alone the species was known from four additional grids. These grids have 20 additional collecting events without any evidence for *M. arthriticus*. Collecting at the old localities is nevertheless urgently needed to confirm the status of the species.

Proposing Red Lists is increasingly important for invertebrates given that the habitat for many species is quickly vanishing. In order to give any credence to such efforts, it is necessary to base ranking decisions on a quantitative assessment of data. As our example demonstrates, in some cases the data are already available, existing deficiencies can be determined and addressed through analytical techniques or additional collecting. In other cases, a compilation of all data will reveal that there is not sufficient information for an assessment. However, compilation of the existing data will at least reveal where the information gaps are. Some taxonomists may argue that guesses by an expert are sufficient, but ultimately guesses are also based on data, and for those insects where identification requires microscopic study these data will come with label information; i.e., one may as well reveal the quantitative data that support the guesses.

## 4. Use of Specimen Data for Estimating Clade Species Richness

Many attempts have been made to estimate the number of species on our planet. Vascular plants and vertebrates are comparatively well-known taxonomically and only relatively few additional species are described every year. The same is not true for fungi and algae, and for invertebrates such as insects, crustaceans and nematodes. Here, new field work produces millions of specimens every year with many of these belonging to undescribed species. The estimates for the total number of species on our planet range between 3–80 million species, with the recent proposals favoring from 5–10 million species (Gaston 1991, Ødegaard 2000, Stork 1988). These estimates are based on a variety of techniques. For example, Erwin (1982) and Ødegaard (2000) used the number of known species of plants and estimates of host-specificity for phytophagous insects. Hodkinson & Casson (1991) compared the number of described to the number of undescribed species in a sample of true bugs (Hemiptera) from Sulawesi. As can be seen from these examples, most studies either only concentrate on a single taxon and/or a single sample from field work in tropical forests. The problems for estimating the global species diversity from such a basis are obvious. Estimates based on a few taxa or samples will be inherently unstable because different taxa have very different ecological requirements that result in different species distributions. For example, as mentioned earlier, Asilidae are most speciose in arid and semiarid environments around the world and have therefore their highest species diversity outside of the equatorial belt.

Specimen data from the thousands of taxonomic revisions can help to overcome the taxon bias in previous estimates of the global species richness. Just imagine having estimates of species richness for the thousands of taxa covered in the thousands of taxonomic revisions that have been published in the last decades. These estimates would cover a wide variety of taxa and geographic regions. In order to derive such estimates, we can again make use of the quantitative information inherent in specimen lists. This information can be used in conjunction with statistical species-richness estimation techniques that were first proposed for estimating the number of unknown species in ecological samples (Colwell & Coddington 1994). These methods are designed to estimate the full species diversity in a sample even if only a subset of the species has been collected. But these techniques can also be applied to specimen data from museums and taxonomic revisions (Meier & Dikow 2004, Petersen *et al.* 2003, Soberón *et al.* 2000).

Here, we present two examples for such species-richness estimations based on specimen lists from taxonomic revisions. One is for the predominantly African robber-fly genus *Euscelidia* Westwood (Dikow 2003)

Table 12.2. Summary of species richness estimation of Afrotropical Asilidae.

| | Afrotropical Region | Republic of South Africa |
|---|---|---|
| Known Species Number | ≈ 1500 | n.a. |
| Species from Revisions | 710 | 470 |
| Specimens from Revisions | 21058 | 15467 |
| Singleton Species | 110 | 76 |
| Extrapolated Species Number | 997-1094 | 668-738 |

and the other for the sub-Saharan robber-fly example used previously. In both cases, we use two non-parametric estimators (ICE – incidence-based estimator and Jack2 – second-order Jackknife, 300 random sample-order runs) as implemented in EstimateS (Colwell 2000). Meier & Dikow (2004) submitted the specimen data from the *Euscelidia* revision to species-richness estimation. The revision recognized 68 species distributed primarily in the Afrotropical Region (55 species) with additional species being found in the Oriental and Palaearctic regions. Overall, some 1,500 specimens had been studied and 14 species are currently only known from the holotype. The non-parametric species-richness estimation techniques indicated that there might still be 36–48 additional, uncollected species (Fig. 12.5a). On the other hand, the fauna of the Republic of South Africa is relatively well-sampled and only 3–5 new species can be expected here (Meier & Dikow 2004). The fauna of the Oriental Region has 11 species, of which 4 are known only from the holotype, and it is therefore the least known fauna and the estimators were unable to suggest an estimate; i.e., too few data points were available for an estimate.

When we apply species-richness estimation techniques to our data set including 710 sub-Saharan robber-fly species with sufficient data for estimation, the results imply that 287–384 species are still waiting to be discovered in sub-Saharan Africa (Fig. 12.5b, Table 12.2). Focusing the attention on the best sampled area, the Republic of South Africa, with 470 species, the number of 198–268 new species is still relatively high (Fig. 12.5c, Table 12.2). This indicates that even after extensive taxonomic work including many targeted field trips conducted by Jason Londt over the past 28 years this highly diverse region is still undersampled. This might not come as a surprise as a single person can hardly be held responsible for revising such a diverse fauna, but it also highlights how far a revisionary effort has come and that more taxonomic work needs to be carried out in order to completely cover the fauna.

But how can we obtain a better idea about the global species richness based on such estimates? Thousands of taxonomic revisions with specimen data can be translated into thousands of point estimates of clade richness for different taxa. Instead of relying on a single taxon for the estimate or relying on only tropical samples, this technique would provide repeatable estimates for many taxa. Given that taxonomic experts tend to revise all species in a taxon and not only the tropical species, we should also include predominantly subtropical and temperate clades in the estimates. Currently, most taxonomic revisions start with discussing the taxonomic history and biology of the revised taxa and then proceed to presenting the main results of the revision. It is here that we believe taxonomists should also estimate the true species richness of the revised clade. Some taxonomists have done so, but for the most part taxonomic revisions lack any quantitative evaluation of the data that have been generated.

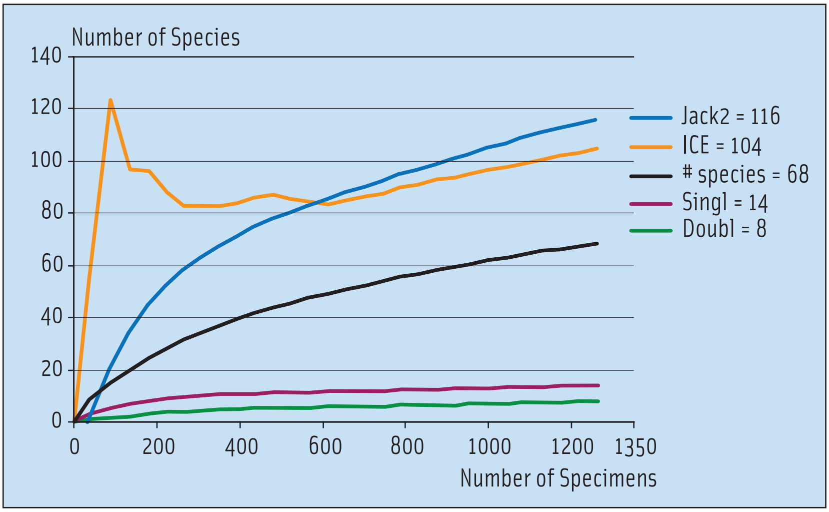


Figure 12.5a. Genus *Euscelidia* Westwood with all species included. Jack2 = second-order Jackknife estimator; ICE = incidence-based estimator; # species = number of observed species; singl = singleton species; doubl = doubleton species.

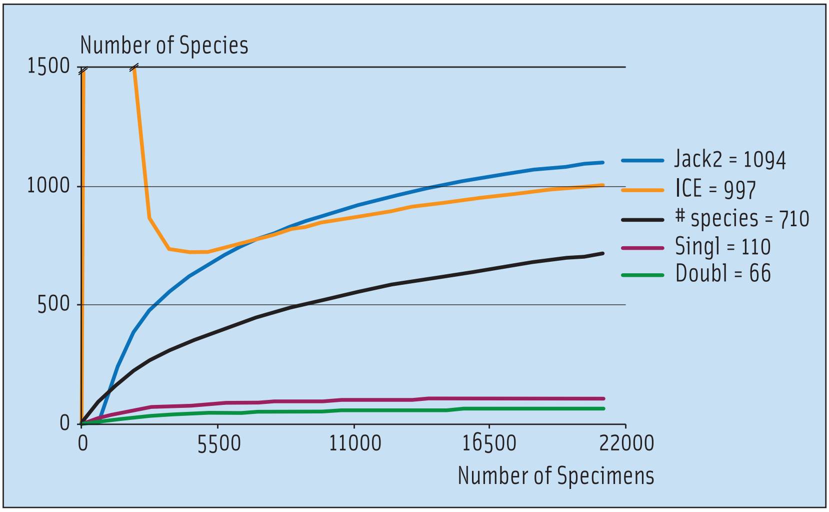


Figure 12.5b. All species of Afrotropical Asilidae data set. Jack2 = second-order Jackknife estimator; ICE = incidence-based estimator; # species = number of observed species; singl = singleton species; doubl = doubleton species.

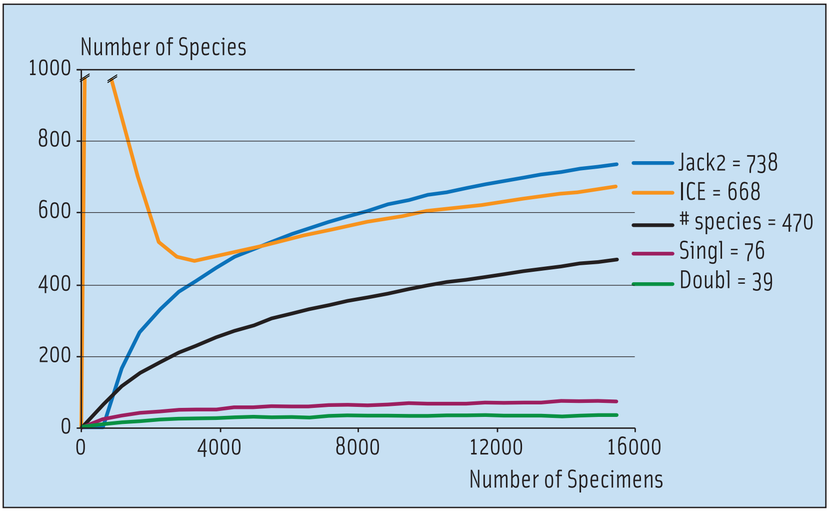


Figure 12.5c. South African species only data set. Jack2 = second-order Jackknife estimator; ICE = incidence-based estimator; # species = number of observed species; singl = singleton species; doubl = doubleton species.

## 5. Use of Specimen Data: The Numerous Problems

So far we have painted an optimistic picture of how specimen lists from taxonomic revisions can potentially provide the answers to long-standing questions. However, the reality is less favorable, or as Ponder *et al.* (2001) have remarked: 'shortcomings of the data include the ad-hoc nature of the collections, presence-only data, biased sampling, and large collecting gaps in time and space …'; i.e., specimen lists from taxonomic revisions have many problems and in many cases the data can mostly serve as a baseline for future work only. The most obvious shortcoming is that the specimen samples used in taxonomy are highly non-random in many ways. Taxonomists tend to concentrate on collecting particular species. For example, rare and conspicuous species are over-represented in collections, while very common species are underrepresented. This bias extends to collecting localities, with promising and easily accessible areas being over-collected and 'difficult' and 'uninteresting' areas being essentially unexplored. For example, combining the known collecting localities for many taxa in the Amazon Basin yields a river map for the Amazon and its main tributaries that can be traversed by boat (Heyer *et al.* 1999), and a map with all localities for Papilionidae and Pieridae butterflies in Mexico resembles this country's highway map (Soberón *et al.* 2000). Collectors furthermore do not record if a particular collecting attempt was unsuccessful; i.e., there are no empty pins with label data in museum collections. All these biases are reflected in the specimen lists in taxonomic revisions. The good news is that these lists are more explicit than many other data used in biodiversity research, i.e., some of these biases can be detected and at least partially be removed by techniques such as rarefaction.

Another mixed blessing of revision data is the large proportion of old specimens. Obviously, old specimens are of great importance in documenting that a species has been present in a particular locality and for documenting that a population has continuously occupied a certain area. However, for present-day conservation decisions they are only of limited value.

The last serious problem that we want to mention here is that not all taxonomic revisions are suitable for the kind of quantitative evaluation that we are advocating. There are those that do not include specimen lists because the editors have insisted on the removal of the data from the manuscript. We hope that such cases will become rare and that specimen lists will at least be preserved as supplementary information on journal homepages. Then there are those revisions, for which the specimen lists are no longer available in electronic format or for which the electronic format is unsuitable for databasing. Both are more serious problems than the reader may think. We encountered these problems when preparing our dataset for the sub-Saharan Asilidae. Printed specimen lists only provide the information available on the specimen label so that for data-poor specimens, the number of fields is very different than for data-rich ones. When label data are printed in a running format like it is typical in taxonomic revisions, it becomes near impossible to automatically capture the data. Lastly, many revisions uncover such a large number of new species based on such a small number of specimens that species-richness estimation and species range mapping becomes impossible, e.g., Grimaldi & Nguyen (1999). However, even in these cases we would advocate that taxonomists summarize the numerical data instead of just predicting that there are a large number of undiscovered species in the group.

## 6. Summary and Conclusions

The specimen information in taxonomic revisions is an outstanding source of information for incorporating hyper-diverse taxa into quantitative biodiversity research. This data source is rich given that more than 2,300 taxonomic revisions have been published between 1990–2002 (Meier & Dikow 2004) and thousands of species descriptions with specimen data have been made available in taxonomic journals.

Here we have described how these data can be used for testing the plant-based biodiversity hotspots with Diptera data. This approach holds considerable promise because it allows for the inclusion of hyper-diverse invertebrate taxa into conservation biology studies. These taxa are otherwise rarely considered. We show that the predatory Asilidae are very speciose in the southern African biodiversity hotspots and thus corroborate the hotspots originally described based on vascular plants. Regardless of the explanation for the observed concordance between Asilidae and plants, our data support initiatives channeling conservation funding into the African biodiversity hotspots (Myers 2003, Reid 1998) and contradicts predictions that restricted-range taxa like invertebrates will be poorly represented by areas selected based on standard indicator taxa (Moore *et al.* 2003, van Jaarsveld *et al.* 1998).

 

The use of explicit specimen data for compiling Red Lists of endangered species promises to become another important application of specimen data from taxonomic revisions. The case of the Danish Asilidae highlights the suitability of this approach for well-sampled regions of the world and corroborates the notion that human development and high population density may increase extinction risks, even for historically common species. The data for the Danish Red List were compiled from a revision of the Danish fauna and the specimens came from two natural history collections. However, instead of just digitizing the label information, all specimens were re-identified in order to reduce misidentification problems.

Clade species-richness estimation based on specimen data is also a method that can easily be applied to many invertebrate taxa and therefore provides a much better estimate of the total species diversity. Adding species richness estimates of a variety of taxa from different biogeographical regions can then be used to extrapolate the number of species roaming on Earth. The result would be a much more accurate estimate for global species diversity because it would be based on multiple, independent data sources and analyzed in a standardized way. The estimation techniques can also direct future collecting efforts to undersampled areas for which most species are only known from a few specimens (Meier & Dikow 2004).

Specimen data from taxonomic revisions also have numerous shortcomings, however. Overcoming these will be important. Fortunately, specimen data are explicit enough that they can be evaluated for collecting biases that can then be addressed by a variety of techniques including targeted new collecting, the use of comparison areas, species-richness estimation, data-set resampling, and the modeling of species distributions (Colwell & Coddington 1994, Meier & Dikow 2004, Williams *et al.* 2002). We would nevertheless like to point out that not all published revisions are suitable for the approaches outlined above. In particular, more standardized formats for all specimen data would improve usefulness of these data.

The three approaches outlined here may help the beleaguered field of revisionary taxonomy (Godfray 2002, Wheeler 2004) by providing a high-impact use for specimen data generated during relatively low-impact taxonomic revisions. Given the potential of the data, conservation biologists and systematists should start collecting specimen data from revisions into a 'specimen bank' of similar design and accessibility as GenBank. Such a database would quickly evolve into the most comprehensive and reliable global source of data on species distributions for a wide variety of taxa (Godfray 2002).

## Acknowledgements

We would like to acknowledge financial support from the European Commission's 'Improving Human Potential' program (Transnational Access to major Research Infrastructures) that supported this research through a grant from the Copenhagen Biosystematics Centre (COBICE) to TD. TD would also like to acknowledge the support from Cornell University and the American Museum of Natural History in form of a graduate student fellowship. Further funding supporting the research came from the Danish Natural Science Research Council (grant nos. 9502155 and 9801904), and FRC grants financed by the Ministry of Education of Singapore (R-377-000-025-112, R-377-000-024-101). We also acknowledge the help from numerous colleagues, who have over the years provided help with carrying out quantitative diversity research on Diptera.

## References

Barthlott, W., Lauer, W. & Placke, A. (1996) Global distribution of species diversity in vascular plants: towards a world map of phytodiversity. *Erdkunde* 50: 317–327.

Brooks, T., Balmford, A., Burgess, N., Fjeldså, J., Hansen, L.A., Moore, J., Rahbek, C. & Williams, P. (2001) Toward a blueprint for conservation in Africa. *BioScience* 51: 613–624.

Brooks, T.M., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Rylands, A.B., Konstant, W.R., Flick, P., Pilgrim, J., Oldfield, S. & Magin, A. (2002) Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology* 16: 909–923.

Colwell, R.K. (2000) EstimateS: *Statistical estimation of species richness and shared species from samples*, Version 5. User's Guide and application. Available at http://viceroy.eeb.uconn.edu/estimates.

Colwell, R.K. & Coddington, J.A. (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 345: 101–118.

Dikow, T. (2003) Revision of the genus *Euscelidia* Westwood, 1850 (Diptera: Asilidae: Leptogastrinae). *African Invertebrates* 44: 1–131.

Erwin, T.L. (1982) Tropical forests: their richness in Coleoptera and other arthropod species. *Coleopterists Bulletin* 36: 74–75.

Fagan, W.F. Kareiva, P.M (1997) Using compiled species lists to make biodiversity comparisons among regions: a test case using Oregon butterflies. *Biological Conservation* 80: 249–259.

Gärdenfors, U. (2001) Classifying threatened species at national versus global levels. *Trends in Ecology and Evolution* 16: 511–516.

Gärdenfors, U., Hilton-Taylor, C., Mace, G.M. & Rodríguez, J.P. (2001) The application of IUCN Red List criteria at regional levels. *Conservation Biology* 15: 1206–1212.

Gaston, K.J (1991) The Magnitude of Global Insect Species Richness. *Conservation Biology* 5: 283–296.

Godfray, H.C. J. (2002) Challenges for taxonomy. *Nature* 417: 17–19.

Grimaldi, D. & Nguyen, T (1999) Monograph on the spittlebug flies, genus *Cladochaeta* (Diptera: Drosophilidae: Cladochaetini). *Bulletin of the American Museum of Natural History* 241: 1–326.

Heyer, W.R., Coddington, J., Kress, W.J., Acevedo, P., Cole, D., Erwin, T.L., Terry, J.B., Meggers, J., Pogue, M.G., Thorington, R.W., Vari, R.P., Weitzman, M.J. & Weitzman, S.H. (1999) Amazonian biotic data and conservation decisions. *Ciência e Cultura São Paulo* 51: 372–385.

Hodkinson, I.D. & Casson, D. (1991) A lesser predilection for bugs: Hemiptera (Insecta) diversity in tropical rain forests. *Biological Journal of the Linnean Society* 43: 101–109.

Humphries, C.J., Vane-Wright, R.I. & Williams, P.H. (1991) Biodiversity Reserves: Setting new priorities for the conservation of wildlife. *Parks* 2: 34–38.

IUCN – The World Conservation Union (2001) IUCN Red List Categories and Criteria, Version 3.1. IUCN Publications Services, Cambridge, 30 pp.

Larsen, N.M. & Meier, R. (2004) The species diversity, distribution, and conservation status of the Asilidae (Diptera: Insecta) in Denmark. *Steenstrupia* 28: 177–241.

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 and complete specimen data. *Conservation Biology* 18: 478–488.

Mittermeier, R.A., Mittermeier, C.G., Brooks, T.M., Pilgrim, J.D., Konstant, W.R., da Fonseca, G.A.B. & Kormos, C. (2003) Wilderness and biodiversity conservation. *Proceedings of the National Academy of Sciences of the United States of America* 100: 10309–10313.

Moore, J.L., Balmford, A., Brooks, T., Burgess, N.D., Hansen, L.A., Rahbek, C. & Williams, P.H. (2003) Performance of sub-Saharan vertebrates as indicator groups for identifying priority areas for conservation. *Conservation Biology* 17: 207–218.

Myers, N. (2003) Biodiversity hotspots revisited. *BioScience* 53: 916–917.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858.

 

Ødegaard, F. (2000) How many species of arthropods? Erwin's estimate revised. *Biological Journal of the Linnean Society* 71: 583–597.

Petersen, F.T., Meier, R. & Larsen, M.N. (2003) Testing species richness estimation methods using museum label data on the Danish Asilidae. *Biodiversity and Conservation* 12: 687–701.

Platnick, N.I. (1992) Patterns of Biodiversity. Pages 15–24 *in*: Eldredge, N. (ed.), *Systematics, Ecology, and the Biodiversity Crisis*. Columbia University Press, New York.

Ponder, W.F., Carter, G.A., Flemons, P. & Chapman, R.R. (2001) Evaluation of museum collection data for use in biodiversity assessment. *Conservation Biology* 15: 648–657.

Reid, W.V. (1998) Biodiversity hotspots. *Trends in Ecology and Evolution* 13: 275–280.

Sechrest, W., Brooks, T.M., da Fonseca, G.A.B., Konstant, W.R., Mittermeier, R.A., Purvis, A., Rylands, A.B. & Gittleman, J.L. (2002) Hotspots and the conservation of evolutionary history. *Proceedings of the National Academy of Sciences of the United States of America* 99: 2067–2071.

Soberón, J.M., Llorente, J.B. & Oñate, L. (2000) The use of specimen-label databases for conservation purposes: An example using Mexican Papilionid and Pierid butterflies. *Biodiversity and Conservation* 9: 1441–1466.

Stork, N.E. (1988) Insect diversity: facts, fiction and speculation. *Biological Journal of the Linnean Society* 35: 321–337.

van Jaarsveld, A.S., Freitag, S., Chown, S.L., Muller, C., Koch, S., Hull, H., Bellamy, C., Kruger, M., Endrödy-Younga, S., M.W. Mansell & Scholz, C.H. (1998) Biodiversity assessment and conservation strategies. *Science* 279: 2106–2108.

Vane-Wright, R.I., Humphries, C.J. & Williams, P.H. (1991) What to Protect Systematics and the Agony of Choice. *Biological Conservation* 55: 235–254.

Wheeler, Q.D. (2004) Taxonomic triage and the poverty of phylogeny. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 359: 571–583.

Williams, P.H., Margules, C.R. & D.W. Hilbert. (2002) Data requirements and data sources for biodiversity priority area selection. *Journal of Biosciences* 27: 327–338.