

---

# Significance of Specimen Databases from Taxonomic Revisions for Estimating and Mapping the Global Species Diversity of Invertebrates and Repatriating Reliable Specimen Data

RUDOLF MEIER\* AND TORSTEN DIKOW†

\*Zoological Museum, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark, and Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, email dbsmr@nus.edu.sg

†Universität Rostock Institut für Biodiversitätsforschung, Allgemeine und Spezielle Zoologie, Uniplatz 2, D-18055 Rostock, Germany, and Cornell University, Department of Entomology, Comstock Hall, Ithaca, NY 14853, U.S.A.

---

**Abstract:** *We argue that the millions of specimen-label records published over the past decades in thousands of taxonomic revisions are a cost-effective source of information of critical importance for incorporating invertebrates into biodiversity research and conservation decisions. More specifically, we demonstrate for a specimen database assembled during a revision of the robber-fly genus Euscelidia (Asilidae, Diptera) how nonparametric species richness estimators (Chao1, incidence-based coverage estimator; second-order jackknife) can be used to (1) estimate global species diversity, (2) direct future collecting to areas that are undersampled and/or likely to be rich in new species, and (3) assess whether the plant-based global biodiversity hotspots of Myers et al. (2000) contain a significant proportion of invertebrates. During the revision of Euscelidia, the number of known species more than doubled, but estimation of species richness revealed that the true diversity of the genus was likely twice as high. The same techniques applied to subsamples of the data indicated that much of the unknown diversity will be found in the Oriental region. Assessing the validity of biodiversity hotspots for invertebrates is a formidable challenge because it is difficult to decide whether species are hotspot endemics, and lists of observed species dramatically underestimate true diversity. Lastly, conservation biologists need a specimen database analogous to GenBank for collecting specimen records. Such a database has a three-fold advantage over information obtained from digitized museum collections: (1) it is shown for Euscelidia that a large proportion of unrevised museum specimens are misidentified; (2) only the specimen lists in revisionary studies cover a wide variety of private and public collections; and (3) obtaining specimen records from revisions is cost-effective.*

Significado de Bases de Datos de Especímenes de Revisiones Taxonómicas para la Estimación y el Mapeo de la Diversidad Global de Especies de Invertebrados y de la Repatriación de Datos Confiables de Especímenes

**Resumen:** *Sostuvimos que los millones de registros de especímenes publicados en miles de revisiones taxonómicas en décadas anteriores son una fuente de información costo-efectiva de importancia crítica para la incorporación de invertebrados en decisiones de investigación y conservación. Más específicamente, para una base de datos de especímenes de moscas del género Euscelidia (Asilidae, Diptera) demostramos como se pueden utilizar estimadores no paramétricos de riqueza de especies (Chao 1, estimador de cobertura basado en incidencia, navaja de segundo orden) para (1) estimar la diversidad global de especies, (2) dirigir colecciones*

---

*Paper submitted May 30, 2002; revised manuscript accepted August 14, 2003.*

futuras a áreas que están sub-muestreadas y/o probablemente tengan especies nuevas y (3) evaluar si los sitios globales de importancia para la biodiversidad basados en plantas de Myers et al. (2000) contienen una proporción significativa de invertebrados. Durante la revisión de *Euscelidia* el número de especies conocidas fue más del doble, pero la estimación de riqueza de especies reveló que la diversidad real del género probablemente también era el doble. Las mismas técnicas aplicadas a las sub-muestras de datos indicaron que gran parte de la diversidad no conocida se encontrará en la Región Oriental. La evaluación de la validez de sitios de importancia para la biodiversidad de invertebrados es un reto formidable porque es difícil decidir si las especies son endémicas de esos sitios y si las listas de especies observadas subestiman dramáticamente la diversidad real. Finalmente, los biólogos de la conservación requieren de una base de datos de especímenes análoga a GenBank, para obtener registros de especímenes. Dicha base de datos tiene una triple ventaja sobre la información obtenida de colecciones de museos digitalizadas. (1) Se muestra para *Euscelidia* que una gran proporción de especímenes de museo no revisados están mal identificados. (2) Sólo las listas de especímenes en estudios de revisión cubren una amplia variedad de colecciones privadas y públicas. (3) La obtención de registros en revisiones es costo-efectiva.

## Introduction

Taxonomic revisions and monographs of animal and plant taxa are the staple of research in systematic biology. For example, a literature search of only one prominent database covering the zoological literature (*Zoological Record*) revealed that more than 2300 revisions and monographs have been published within the last 10 years, and Gaston (1991) found, also using the *Zoological Record*, that within the four hyperdiverse insect "orders" alone more than 10,000 new species were described between 1986 and 1989.

In many animal and plant groups it is standard practice that for a taxonomic revision all museum specimens for the targeted taxon are borrowed from the world's natural history collections. The data on the specimen labels is routinely recorded by taxonomists for generating locality lists and/or distribution maps. We demonstrate for the robber-fly genus *Euscelidia* (Asilidae, Diptera) that such specimen databases have many additional uses that remain largely unexplored.

Imagine that conservation biologists built a database containing all the published specimen data from the thousands of published revisions and monographs of the past 50 years. This database would be very useful for addressing many important issues in quantitative biodiversity and conservation research. Not only could these issues be studied based on a truly impressive amount of data that could be analyzed quantitatively, but much of the data would also come from poorly known taxa such as invertebrates. Furthermore, gathering these data would be relatively inexpensive because the label information would be published and often available in an electronic format. Moreover, specimen identifications would be accurate and dubious localities would have been resolved and often already assigned coordinates.

We discuss the use of specimen data for (1) estimating global species diversity, (2) directing future collecting,

(3) assessing biodiversity hotspots for invertebrates (Myers et al. 2000), and (4) obtaining reliable specimen information for data repatriation to the country of specimen origin.

Over the years, many answers to the question of how many species are on our planet have been attempted, but agreement remains elusive because different estimation approaches yield wildly differing results ranging from 3 to 80 million, with most recent work favoring 5–10 million species (e.g., Stork 1988; Gaston 1991; Hodkinson & Casson 1991; Ødegaard 2000). One approach relies on the proportion of undescribed to described species in a particular sample or taxon. For example, in Sulawesi Hodkinson and Casson (1991) collected 1690 species of Hemiptera, of which 62.5% were undescribed. Given that 81,700 species of terrestrial Hemiptera were known in 1990 and under the assumption that the same proportion of Hemiptera remains undescribed worldwide as there was in the Sulawesi sample, the global species estimate would be roughly 180,000. If every tenth insect species belongs to the Hemiptera, a global species estimate for insects would be around 1.8 million species.

However, an estimate based on a single proportion of undescribed species in one sample is hardly satisfactory. The correct proportion will differ from taxon to taxon and from region to region. It is here that the specimen data from taxonomic revisions can make a valuable contribution. Each revision can yield a point estimate of this proportion by comparing the numbers of described species before and after the revision.

This estimate would still only pertain to the number of species available in the natural history collections around the world (collected species). There will almost always be additional species that remain uncollected, and estimating the proportions of these uncollected species is essential for obtaining an accurate picture of global species diversity. Such an estimate can be attempted by using the species-richness estimation techniques developed

during the past decades (review in Colwell & Coddington 1994). Although these techniques were initially proposed for samples obtained under standardized sampling protocols in ecology, researchers have recently also started to apply them to museum-collection samples (e.g., Heyer et al. 1999; Soberón et al. 2000; Petersen et al. 2003). Species estimators have two advantages over the traditionally used species-description and species-accumulation curves. First, they use information about the proportion of rare species in a sample to judge its completeness and attempt an extrapolation from the observed number of species to the true species diversity of the sampling universe. Second, estimator curves often flatten well before the corresponding species-accumulation curves, allowing for a species-diversity estimate at sample sizes at which species-accumulation curves continue to rise and thus fail to suggest a final value. We demonstrate for *Euscelidia* how these techniques can be used to estimate the proportion of uncollected species.

There is general agreement that biodiversity research and conservation decisions should not be based only on relatively well-known taxa such as birds and large mammals but also on invertebrates (e.g., Myers et al., 2000). Currently the lack of invertebrate data effectively prevents conservation biologists from increasing their taxonomic coverage in evaluating, monitoring, and choosing reserves. One obvious way to incorporate invertebrates would be through massive collecting in the areas under consideration for conservation. Funding for collecting is limited, however, and the frequently large numbers of rare species (e.g., Novotný & Basset 2000) would require unrealistically large samples before good coverage could be achieved.

The next-best solution would be to combine the specimen information that has accumulated over hundreds of years in collections and the systematic literature with information from collecting activities that target the critically undersampled areas. A quantitative evaluation of the specimen lists from taxonomic revisions would constitute the first step in this process. Preliminary distributions for the revised species can be plotted and the quality of the taxon sampling can be assessed by using species-accumulation curves and applying species-richness estimation techniques. The approach is essentially the same as described for estimating global species diversity. For any region of interest, the proportion of uncollected species can be estimated based on species richness estimation (cf. "completeness ratio" of Soberón et al. 2000). Information on the sampling coverage of different regions can afterward be used to justify further collecting of a particular taxon in a particular region, or a large number of revisions can be screened for those areas that are repeatedly found to be undersampled. The latter information would be particularly important for organizations such as the All-Species Foundation, which could then

target these poorly collected regions for general-survey expeditions.

The 25 biodiversity hotspots of Myers et al. (2000) comprise only 1.4% of the entire land surface, yet today 44% of all species of vascular plants and 35% of all species in four vertebrate groups are confined to these areas. However, we do not know how well these hotspots hold up for invertebrates. Important and relatively easily accessible data are again hidden in the specimen databases of taxonomic revisions. With a database covering thousands of revisions at hand, one could map the known distributions of many species based on a large amount of specimen-label data. One could determine the number of species that are either restricted to or found within hotspots. The data could furthermore be mined for information on other issues, such as the species complementarity of invertebrates at different sites (Colwell & Coddington 1994; Bartlett et al. 1999). All this work would not have to rely on lists of observed species because, with the availability of specimen-label data, researchers would be able to conduct species-richness estimates within hotspots. We demonstrate the power of this approach through our *Euscelidia* example.

Some people claim that the natural history museums around the world contain "the most comprehensive, reliable source of knowledge for most described species. . ." (Ponder et al. 2001; see also Gaston et al. 1995). In large museums, much of the data pertains to specimens not collected in the home country of the institution (e.g., approximately 50% of the Copenhagen Diptera collection). One goal of many recent biodiversity initiatives (e.g., Global Biodiversity Information Facility 2000) is repatriating the specimen data. Two approaches are conceivable. The more popular one relies on "digitizing" the specimen-label information in the natural history collections. An alternative approach is to create a specimen bank analogous to GenBank for specimen data published in taxonomic revisions (cf. Godfray 2002a, 2002b). We demonstrate for *Euscelidia* that the latter approach has several advantages with regard to the quality of the specimen information, the specimen coverage, and the cost of obtaining the data.

As our example, we chose a data set for a group of predatory flies (*Euscelidia*: Asilidae: Diptera) and used a specimen database generated during a revision of *Euscelidia* (Dikow 2003). The Asilidae are one of the largest families of Diptera, containing 6900 described species. *Euscelidia* is a relatively large genus and predominantly occurs in the Afrotropical region (55 species). Additional species are found in the Oriental (11 species) and the Palearctic region (4 species). Prior to the revision, 29 species had been described. A revision of 1383 specimens from 19 collections revealed 40 new species and placed 4 species in synonymy (Dikow 2003). Species of *Euscelidia* live in grass-dominated habitats, such as grasslands, acacia savannas, or the margins of forest and bush

land. Generally, the specimens in natural history collections have been collected by sweeping, but some have also been caught in Malaise traps. Both asilid specialists and nonspecialist collectors contributed to the sample, but *Euscelidia* was never targeted directly.

## Methods

To obtain an estimate of the number of taxonomic revisions and monographs in zoology published after 1990 (1990–2002), we conducted a search in the *Zoological Record* online database for the words *revision* or *monograph* and discarded references that used the search terms in a nontaxonomic context, had unclear titles, or revised only a single species.

We assembled a database for all 1383 examined specimens of *Euscelidia* by recording species name, zoogeographical region, gender, locality (country, coordinates), collecting date, collector, and depository. We determined the proportion of misidentified, identified, and unidentified specimens by consulting loan forms and notes. To illustrate taxonomic progress on *Euscelidia*, we prepared a cumulative description curve and supplemented it with a cumulative collection curve plotting the year in which a species was collected for the first time against the cumulative number of species (cf. Bickel 1999). We plotted this curve from 1893 onward because most old specimens lack collecting dates.

We used this database to estimate species richness with EstimateS (Version 6.0b1; Colwell 2000) and 300 random sample-order runs. Five-year collecting periods were used as subsampling schemes, and specimens from the nineteenth century lacking label data were considered collected in the year of species description. We estimated the richness for the following areas: global fauna, sub-Saharan Africa without South Africa, South Africa, Oriental region, and three biodiversity hotspots. Only 48 out of 50 species were incorporated in the sub-Saharan Africa estimate because there was no label data for the sole specimen of *E. longibifida* and no voucher specimen for *E. nitida*. We plotted the species-accumulation curve, Coleman curve, number of singletons and doubletons, one abundance-based estimator (Chao1), one incidence-based estimator (ICE), and the second-order jackknife estimator against the number of specimens.

We used ArcView (version 3.1) to plot biodiversity hotspots as recognized by Myers et al. (2000) on a background map containing country borders. The electronic shape files of the hotspots were obtained from Conservation International (2001). We counted the number of species of *Euscelidia* that occurred in or that were endemic to the different hotspots, recorded the number of collecting events in the hotspots, and carried out species richness estimates for the western African forests, Eastern

Arc and Coastal Forests of Tanzania/Kenya (hereafter abbreviated to “Eastern Arc”), and Western Ghats/Sri Lanka.

## Results

Our search resulted in 3983 hits, of which 1455 were revisions of genera or species groups on a global scale and 925 on a regional scale.

### Evaluation of the Revisionary Data

Of the 1383 specimens included in this study, 361 (26%) were identified prior to the revision. Of these, 83% were incorrectly identified either due to misidentification (73%) or synonymy (10%). In one case, however, 101 specimens with identical locality information had been misidentified, and when we counted this case only as a single misidentification the proportions were 62% and 13%, respectively. The cumulative description curve was characterized by a monotonous and slow increase in descriptions until 1950 (Fig. 1) and by two sharp jumps in the period from 1953 to 1957 and in 2002 (as a result of work by Janssens [1953, 1954a, 1954b, 1957] and Dikow [2003]). The collection curve indicated that new species of *Euscelidia* accumulated quickly from 1900 onward, with no evidence of a marked slowdown in recent years (Fig. 1). Fifty-seven percent of all species were unknown prior to the revision. To obtain this proportion, we divided the difference between the number of known species prior to revision (29 spp.) and after revision (68 spp.) by the latter.

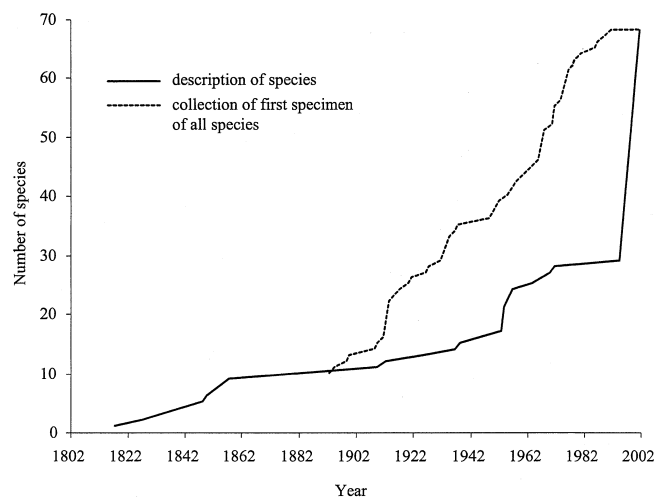
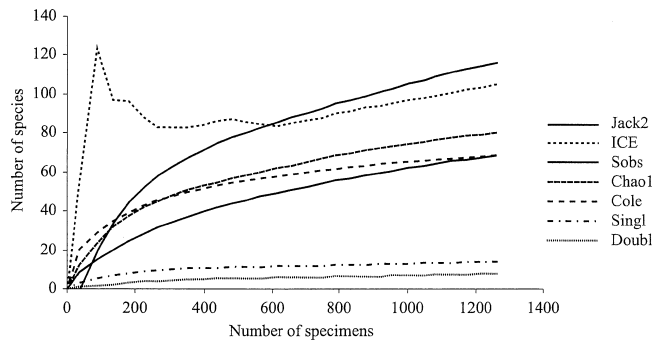


Figure 1. Cumulative species description and species-collection curve for *Euscelidia*. The collection curve starts in 1893 because earlier specimens generally lack labels with year of collection.



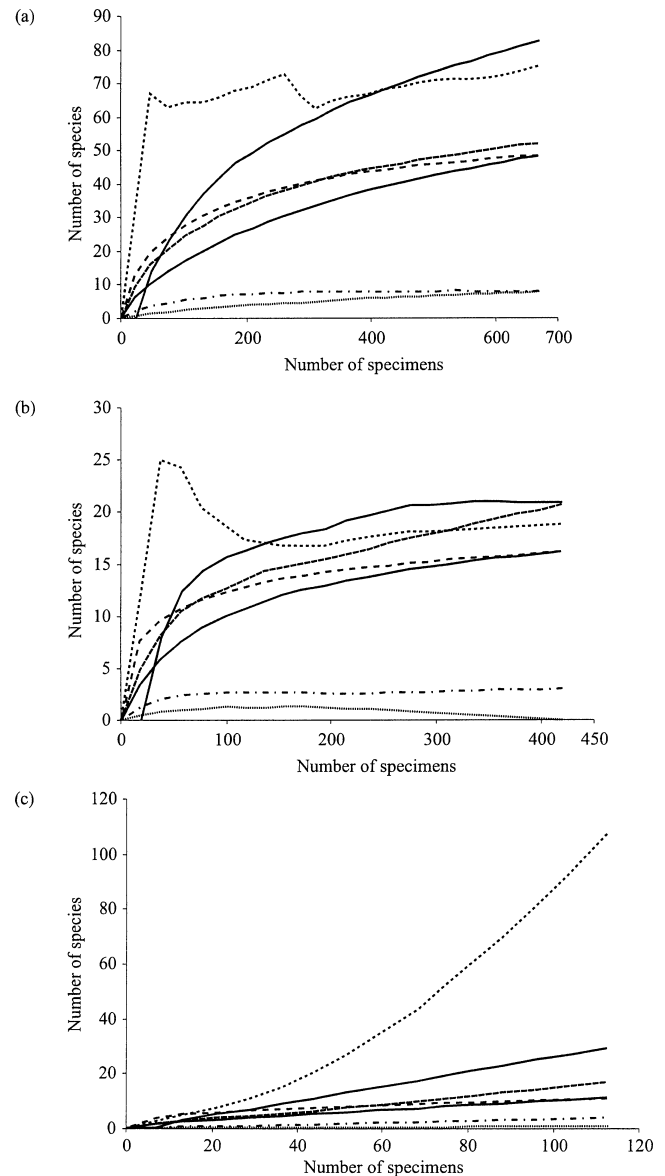
**Figure 2.** Species-richness estimation for *Euscelidia*. Abbreviations: Jack2, second-order jackknife; ICE, incidence-based coverage estimator; Sobs, observed number of species; Chao1, Chao1 estimator; Cole, Coleman curve; Singl, number of singletons; Doubl, number of doubletons.

### Diversity Estimation

None of the species-accumulation curves reached a perfectly satisfactory plateau ("sobs," Figs. 2–4). With the exception of the estimate for the Oriental region, ICE rose quickly, overshoot the final estimate, and then reached a more (e.g., Fig. 3b) or less (e.g., Fig. 2) stable plateau. The abundance-based Chao1 followed the species-estimation curve closely and failed to reach a plateau in all cases (Figs. 2–3c). It usually also gave lower estimates than ICE (Table 1). All curves for the Oriental species failed to show any sign of saturation (Table 1; Fig. 3c). Based on these figures, we computed the proportion of uncollected species for *Euscelidia* as the difference between the estimated and observed richness divided by the estimated richness. For ICE this proportion was 35%, for the second-order jackknife 41%, and for Chao1 15%.

### Hotspot Analysis

Of the 68 known species, 24 (35%) occurred in eight of the biodiversity hotspots of Myers et al. (2000) and 13 were potentially endemic (19%) (Table 2; Fig. 5; Appendix 1). The number of collecting events in the different hotspots was generally low (Table 2). We nevertheless attempted an estimate for the western African forests (Fig. 4a), the Eastern Arc (Fig. 4b), and Western Ghats/Sri Lanka. Because of low species diversity (1–2 species), we did not attempt estimates for the Cape Floral Province and the Mediterranean Basin. For the Western African forests, the estimates indicated at least an additional 5–10 over the observed 7 species and for the Eastern Arc an additional 3–4 over the observed 4 species. The graphs for Western Ghats/Sri Lanka resembled the ones for the Oriental region (Fig. 3c) in that they never reached a plateau.



**Figure 3.** Species-richness estimation for *Euscelidia* occurring in (a) sub-Saharan Africa, excluding the Republic of South Africa, (b) the Republic of South Africa, and (c) the Oriental region. Abbreviations as in Fig. 2.

## Discussion

### Estimation of Global Species Diversity

Before *Euscelidia* was revised, any global species richness estimate would have been guesswork outside the realm of science. But after revision and with the use of species richness estimation, we are now two decisive steps closer to obtaining an appropriate, quantitative estimate. Traditionally, systematists have attempted such estimates based on species description or collection curves (e.g., Steyskal 1965; Dolphin & Quicke 2001). For *Euscelidia*, however,

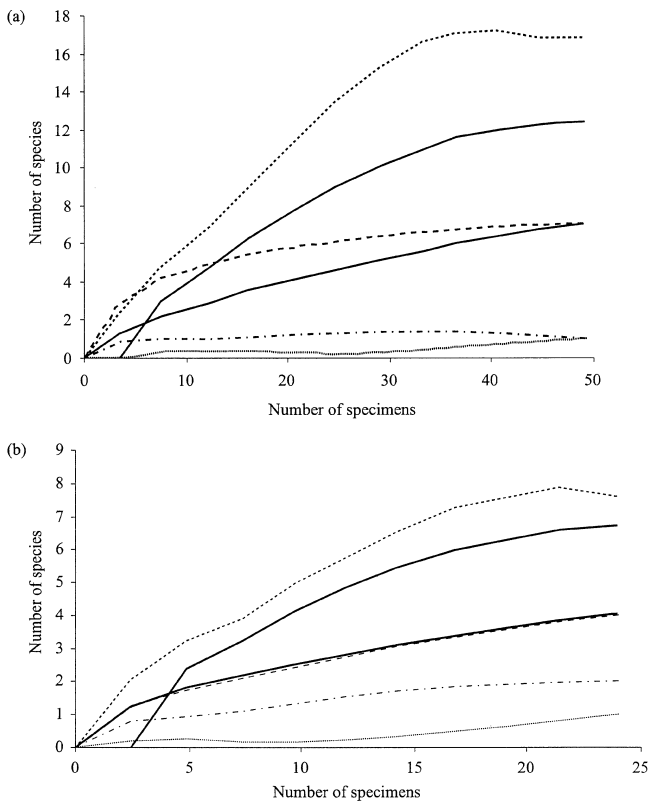


Figure 4. Species-richness estimation for *Euscelidia* occurring in the biodiversity hotspot of (a) western African forests and (b) Eastern Arc. Abbreviations as in Fig. 2.

both curves rise steeply, and the species-description curve (Fig. 1) especially shows the typical periodic bursts of taxonomic activity that make any extrapolation impossible (e.g., Steyskal 1965; McAlpine 1994; Bickel 1999; Dolphin & Quicke 2001). These problems are not shared by the modern species-estimation techniques that utilize information on the abundance of specimens and the proportions of rare species (cf. McAlpine 1994).

Even using these estimators, however, extracting a global estimate for *Euscelidia* is not straightforward be-

cause different estimators come to conflicting conclusions (Table 1) and neither of the estimator curves reaches a perfectly satisfactory plateau. This is not surprising because the same is observed for many samples obtained in ecological studies (e.g., Coddington et al. 1996; McKamey 1999; Anderson & Ashe 2000) or from museum collections (e.g., Heyer et al. 1999). However, it raises the question of which estimate should be used. We argue that Chao1 is an abundance-based estimator and thus unlikely to perform well for taxa like *Euscelidia* that are known to contain species with clumped distributions. We therefore prefer incidence-based estimators.

Based on these estimators *Euscelidia* is expected to have at least 104–116 species. The proportion of the uncollected fauna is thus between 36% and 41%. This implies that unrevised groups similar to *Euscelidia* might have as many as 3.7–4 times the number of species that are currently described (57% undescribed + 36–41% uncollected). One might wonder whether such a high estimate is realistic, but it is in line with other recent Diptera revisions (e.g., Londt 1985, 1988; McAlpine 1994; Grimaldi & Nguyen 1999), with a recent estimate for Braconidae (Hymenoptera; Dolphin & Quicke 2001), and with many of the expert assessments for hyperdiverse insect “orders” reported by Gaston (1991). Furthermore, the proportion of singleton species in our data (20%) is high, indicating that our *Euscelidia* sample is still incomplete.

Regardless of which estimator is used, one should remember that these estimates are for several reasons only approximate. First, estimations are inherently problematic if only approximately half the species in a sample are known (e.g., Colwell & Coddington 1994; Hammond 1994; Dolphin & Quicke 2001). Second, all numbers provided by nonparametric estimators are lower-bound estimates (Colwell & Coddington 1994; Petersen et al. 2003; but see Hellmann & Fowler 1999). Third, the estimators were initially developed for ecological samples, and it is unclear to which degree museum collections based on uneven sampling violate the statistical assumptions of these techniques and how the violation affects the results (Petersen et al. 2003).

Table 1. Estimates of species diversity for *Euscelidia*.\*

Region	Samples	Chao1	ICE	Jack2	Bootstrap	Observed		
						species	Singletons	Doubletons
Entire distribution	29	80 ± 9	104	116	80	68	14	8
Sub-Saharan Africa except Republic of South Africa (RSA)	25	52 ± 4	74	82	56	48	8	8
RSA	21	21 ± 0	19	21	18	16	3	0
Oriental region	15	16 ± 12	107	29	15	11	4	1
Western African forests	12	7 ± 1	17	12	9	7	1	1
Eastern Arc and Coastal Forests of Tanzania and Kenya	10	5 ± 4	8	7	5	4	2	1

\*Estimated numbers and standard deviations are rounded to the next species; the latter are 0 for all estimates based on the incidence-based coverage estimator (ICE), second-order jackknife (jack2), and bootstrap estimates (bootstrap).

**Table 2.** Number of *Euscelidia* species occurring in and potentially endemic to biodiversity hotspots.

Hotspot	Species	Endemic	Collecting events	Localities
Western African forests	7*	1	20	18
Cape Floral Province	2	1	19	14
Eastern Arc	4*	0	16	10
Indo-Burma	3	3	3	3
Madagascar	1	1	1	1
Mediterranean basin	1	0	27	5
Philippines	1	1	1	1
Western Ghats/ Sri Lanka	7	6	21	20
Total	24	13	108	72
Proportion (%)	35.2	19.1		

\*Two species occur in both biodiversity hotspots.

Recent application of species estimation to museum-specimen databases reveals that estimation curves based on collection data behave similarly to those based on ecological samples (e.g., Soberon & Llorente 1993; Kress et al. 1998; León-Cortés et al. 1998; Heyer et al. 1999; Soberón et al. 2000; Petersen et al. 2003), and the only noticeable difference is that the larger samples derived from collections and revisions often yield curves that reach a satisfactory plateau (e.g., Petersen et al. 2003), whereas ecological samples are plagued by large proportions of

rare species (e.g., Novotný & Basset 2000). However, additional work is needed to test how sensitive species richness estimation is to violations of its underlying statistical assumptions. For example, one could test the estimate for a particular site based on the museum-specimen sample against the results obtained in a thorough ecological survey of the same area.

Most approaches to estimating global species diversity emphasize the importance of extrapolating from the known richness of particular field sites to the global scale (e.g., Erwin 1982; 1983; Stork 1988; Hodkinson & Casson 1991; Bartlett et al. 1999; Godfray et al. 1999). Our approach is complementary in that it uses all available specimen information on a particular taxon for extrapolation. It is the first to use species richness estimation techniques and allows for taxon-specific extrapolation. Our approach is closely related to the "all-biota taxon inventory strategy for biodiversity studies" (Wheeler 1995; Platnick 1999), whereas alternative techniques are more closely related to the "all taxa biodiversity inventory" initiative. Both approaches are largely independent and thus have the potential to yield independent estimates for the same value.

#### Identifying Collecting Priorities

Williams et al. (2002) distinguish two steps in identifying biodiversity priority areas. The first involves the collection of good data on distribution and abundance of the

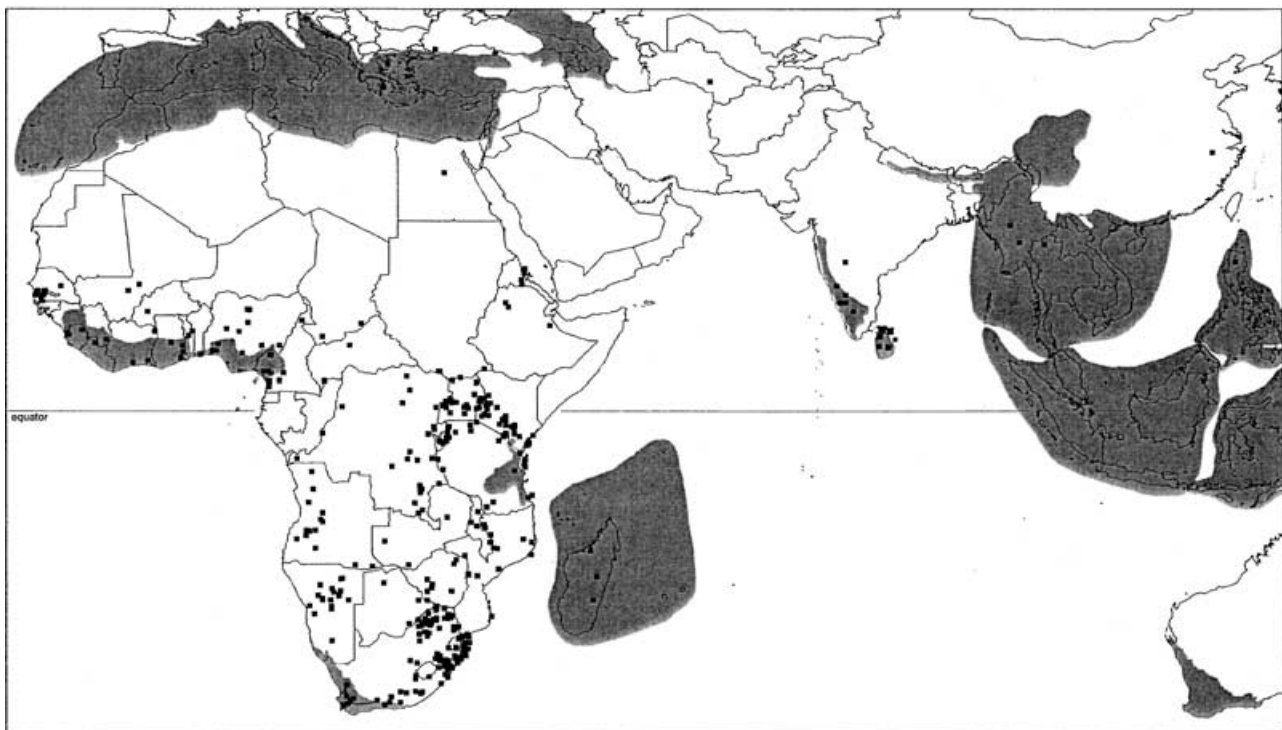


Figure 5. Distribution map for *Euscelidia* with biodiversity hotspots marked in gray (scale: 1:61,000,000).

features to be conserved. We believe that obtaining specimen information from taxonomic revisions is of critical importance in obtaining these data. In many cases, however, they have to be complemented with new collection records. It is thus of considerable importance for conservation biology and systematics to develop quantitative techniques for determining whether additional collecting is necessary and for targeting insufficiently sampled areas during future collecting. It is one of the great advantages of specimen data in revisions that it can be subjected to quantitative study through species richness estimation. The observed species richness can be compared to the predicted species diversity, and sampling can be initiated to correct for biases. For *Euscelidia* the results are clear-cut, and—all other factors being equal—the Oriental fauna should be targeted in future fieldwork because neither the species-accumulation curve nor the estimators show any sign of flattening. In the Afrotropical and the Palearctic regions, the proportion of uncollected *Euscelidia* species is lower. This is particularly so for South Africa, whereas the rest of sub-Saharan Africa remains relatively poorly sampled. The relatively good coverage of South Africa is not surprising because its insect fauna is generally better known than the fauna of remaining Africa and because it is home to one of the world's lead experts in Asilidae (J. Londt).

#### Assessing Biodiversity Hotspots for Invertebrates

Forty-four percent of vascular plants and 35% of all species in four vertebrate groups are endemic to the biodiversity hotspots of Myers et al. (2000). One of the main challenges of conservation biology is to establish whether similar levels of endemism are found in invertebrates. At first glance it appears as if the hotspots are failing for *Euscelidia* (endemism of 19%). But especially in invertebrates, relying on observed species lists “without reference to a taxon sampling-curve is problematic at best” (Gotelli & Colwell 2001). For example, there are currently only a small number of Oriental hotspot endemics in *Euscelidia*, but the species estimate for western Ghats/Sri Lanka alone indicates that, because of undersampling, the number will likely increase dramatically above the current level (9% and 4%; Table 2). The species estimates for two additional hotspots in Africa (Fig. 4) also indicate that the observed number of species is clearly an underestimate and that the correct species numbers are probably twice as high.

Establishing the number of hotspot endemics for invertebrates with any certainty will be very difficult. Invertebrates are not well enough sampled to indicate whether any particular species is really endemic to a particular hotspot. Furthermore, as *Euscelidia* demonstrates, one cannot rely on observed species lists, and species richness estimates are not only imprecise but the “estimated” species also remain anonymous. At most one can prob-

ably obtain reasonable estimates for the total number of species (endemic and nonendemic) living in any particular hotspot and then compare this number to the corresponding figures for plants and vertebrates. For example, the Western African forest is home to an observed 7 (10% of known diversity) and an estimated 12–17 *Euscelidia* species (10–16% of ICE and jackknife estimates). The corresponding figures for vascular plants are 9000 (3%) and for higher vertebrates 1320 (4.8%).

Probably more interesting than figures for individual hotspots are estimates for the “global” proportions of species occurring in all 25 hotspots. Twenty-four (35%) of the 68 *Euscelidia* species occur in at least one hotspot. Unfortunately, these are again only the numbers of observed species, and ideally they should be corrected by adding the number of yet unknown species.

Two different approaches to estimating the expected species richness are conceivable. One would be based on estimating the species distributions for each species through modeling techniques (e.g., Williams et al. 2002). The number of species for a given area could then be determined by counting the number of expected species. The advantage of this approach is that for each area definite species lists could be generated. However, the various models for predicting distributions are largely untested for invertebrates. The alternative approach is to use species-estimation techniques to not only estimate the number of expected species for one area but also to estimate the expected species overlap. Such techniques have been described (Chen et al. 1995) and are implemented in EstimateS (Colwell 2000). Here the identity of the species in the estimated species overlap remains unknown, but this approach has the advantage of not relying on models for predicting species distributions. Future studies need to test which approach is more promising for invertebrates.

#### Repatriation of Specimen Data

There is general agreement that biodiversity data should be repatriated to the country of specimen origin. But where should these data come from? Traditionally, museum-collection digitization has been promoted. However, we argue that for several reasons data from taxonomic revisions are a more promising source.

The first reason is the widespread misidentifications, synonymy, and use of invalid names in museum collections. During the revision of *Euscelidia*, a frightening proportion of the borrowed “determined” material was found to be misidentified (62–73%), and a literature search in a *BIOSIS Previews* revealed that the problem is widespread. For example, of the 1522 rove beetle specimens (Staphylinidae: Coleoptera) in the Struve collection 262 (17%) were misidentified (Rose 2000), and Papp (1978) reports that for a collection of Hungarian Lauxaniidae (Diptera) 28 of the 74 species determined and



labeled by Szilády were consistently misidentified. Another problem is the widespread use of invalid names. For example, in *Euscelidia* 13% of all borrowed specimens were classified under an incorrect name, and for a recent inventory of palm collections in botanical gardens, 260 (22%) of the submitted 1208 names were synonyms and 46 (4%) were invalid (Maunder et al. 2001).

Obviously, misidentifications and synonyms are automatically corrected in revisionary studies and only pose a threat to museum-specimen recording schemes. Here, large-scale re-identification of the holdings is usually not an option given the small number of curators and the high degree of taxon specialization typical for modern systematics.

The second reason to prefer specimen data from revisions over those from natural history inventories is collection coverage. For the *Euscelidia* revision, specimens from 19 museum and private collections were studied. Even if the large-scale digitization of museum collections were to start tomorrow it would take decades for most of these collections to be covered; hence, *Euscelidia* distributions based on museum inventories would be woefully incomplete compared to distributions based on data from the taxonomic revision.

Last, obtaining specimen information from revisionary studies is much more cost-effective than getting similar data through the digitization of museum collections. Many taxonomists keep specimen inventories in an electronic form. Even for published revisions, where the data are no longer available in an electronic format, re-entering the information requires less time than entering label data from specimens. Furthermore, deciphering old label information often requires the kind of knowledge restricted to taxonomic experts in the taxa under revision.

Thus, we believe that more attention and funds should be devoted to collecting the specimen data from taxonomic revisions. Such data are, on average, of high quality, cover the specimens in a large number of collections, and cost comparatively little to collect. Ultimately, the availability of the data depends on funding revisionary systematics, and, although its importance is generally acknowledged (Gaston 1991; Stork 1993; Scoble et al. 1995), this line of research has suffered a tremendous decline (e.g., Wheeler 1990; Smith 2001). It is important, however, to remember that a global species-diversity estimate for *Euscelidia* and other invertebrate taxa would be impossible without prior taxonomic revision, and that plotting species distributions based on misidentified museum specimens will be misleading.

## Conclusions

Specimen records in revisionary studies have long been considered of minor value and are usually either printed

in small font or even deleted by journal editors. As we demonstrate for the data from a single revision, however, they contain a wealth of information and constitute an excellent source for research in conservation biology. Collecting this data and thus getting access to biodiversity information on "the other 99%" of species diversity (Ponder & Lunney 1999) is ultimately dependent on the health of revisionary taxonomy. The techniques we discussed worked fairly well for *Euscelidia*, which is a typical example of a predominantly tropical insect taxon that is neither so hyperdiverse nor so badly undersampled that any attempt at estimating its global diversity is hopeless. There are, however, still many taxa that even after revision will be so poorly sampled that any attempt at understanding their diversity and distribution will fail. For such taxa we have to concur with May (1988) that "although species richness is a natural measure of biodiversity, it is an elusive quantity to measure properly."

## Acknowledgments

We acknowledge helpful discussions and comments on the manuscript by N. Scharff, N. M. Andersen, L. Sørensen, Q. Wheeler, T. Brooks, two reviewers, and the editors of *Conservation Biology*. This research was inspired by Conservation International's initiative on Expanding the Capacity to Map Invertebrates at a Global Scale. Conservation International also provided the shape file for the biodiversity hotspots. T.D. acknowledges funding from the European Commission's program on Improving Human Potential (Transnational Access to Major Research Infrastructures), which supported this research through a grant from the Copenhagen Biosystematics Centre (COBICE). R.M. acknowledges support from the Danish National Science Research Council (9801904 and 9801955) and Academic Research Fund grants from the National University of Singapore (R-377-000-025-112 and R-377-000-024-101).

## Literature Cited

- Anderson, R. S., and J. S. Ashe. 2000. Leaf litter inhabiting beetles as surrogates for establishing priorities for conservation of selected tropical montane cloud forests in Honduras, Central America (Coleoptera: Staphylinidae, Curculionidae). *Biodiversity and Conservation* 9:617-653.
- Bartlett, R., J. Pickering, I. Gauld, and D. Winsor. 1999. Estimating global biodiversity: tropical beetles and wasps send different signals. *Ecological Entomology* 24:118-121.
- Bickel, D. J. 1999. What museum collections can reveal about species accumulation, richness, and rarity: an example from the Diptera. Pages 174-181 in W. Ponder and D. Lunney, editors. *The other 99%: the conservation and biodiversity of invertebrates*. Royal Zoological Society of New South Wales, Mosman, Australia.
- Chen, Y.-C., W.-H. Hwang, A. Chao, and C.-Y. Kuo. 1995. Estimating the number of common species: analysis of the number of common bird species in Ke-Yar Stream and Chung-Kang Stream. *Journal of the Chinese Statistical Association* 33:373-393.

- Coddington, J. A., L. H. Young, and F. A. Coyle. 1996. Estimating spider species richness in a southern Appalachian cove hardwood forest. *Journal of Arachnology* 24:111-128.
- Colwell, R. K. 2000. EstimateS: statistical estimation of species richness and shared species from samples. Version 6.0b1. University of Connecticut, Storrs. Computer software and user's guide distributed by the author; available from <http://viceroy.eeb.uconn.edu/estimates> (accessed April 2003).
- Colwell, R. K., and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London Series B* 345:101-118.
- Conservation International (CI). 2001. Biodiversity hotspots. CI, Washington, D.C.
- Dikow, T. 2003. Revision of the genus *Euscelidia* Westwood, 1850 (Diptera: Asilidae: Leptogastrinae). *African Invertebrates*: 44:1-131.
- Dolphin, K., and D. L. J. Quicke. 2001. Estimating the global species richness of an incompletely described taxon: an example using parasitoid wasps (Hymenoptera: Braconidae). *Biological Journal of the Linnean Society* 73:279-286.
- Erwin, T. L. 1982. Tropical forests: their richness in Coleoptera and other arthropod species. *The Coleopterists Bulletin* 36:74-75.
- Erwin, T. L. 1983. Tropical forest canopies: the last biotic frontier. *Bulletin of the Entomological Society of America* 30:14-19.
- Gaston, K. J. 1991. The magnitude of global insect species richness. *Conservation Biology* 5:283-296.
- Gaston, K. J., M. J. Scoble, and A. Crook. 1995. Patterns of species description: a case study using the Geometridae (Lepidoptera). *Biological Journal of the Linnean Society* 55:225-237.
- Global Biodiversity Information Facility (GBIF). 2000. Business plan for the Global Biodiversity Information Facility. Discussion draft version 5. GBIF Secretariat, Copenhagen. Available from <http://www.gbif.org/gbif.org/facility> (accessed April 2003).
- Godfray, H. C. J. 2002a. How might more systematics be funded? *Antenna*, Bulletin of the Royal Entomological Society 26:11-17.
- Godfray, H. C. J. 2002b. Challenges for taxonomy? *Nature* 417:17-19.
- Godfray, H. C. J., O. T. Lewis, and J. Memmott. 1999. Studying insect diversity in the tropics. *Philosophical Transactions of the Royal Society of London Series B* 354:1811-1824.
- Gotelli, N. J., and R. K. Colwell. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4:379-391.
- Grimaldi, D., and T. Nguyen. 1999. Monograph on the spittlebug flies, genus *Cladochaeta* (Diptera: Drosophilidae: Cladochaetini). *Bulletin of the American Museum of Natural History* 241:1-326.
- Hammond, P. M. 1994. Practical approaches to the estimation of the extent of biodiversity in speciose groups. *Philosophical Transactions of the Royal Society of London Series B* 345:119-136.
- Hellmann, J. J., and G. W. Fowler. 1999. Bias, precision, and accuracy of four measures of species richness. *Ecological Applications* 9:824-834.
- Heyer, W. R., J. A. Coddington, J. W. Kress, P. Acevedo, D. Cole, T. L. Erwin, J. Meggers, M. G. Pogue, R. W. Thorington, R. P. Vari, M. J. Weitzman, and S. H. Weitzman. 1999. Amazonian biotic data and conservation decisions. *Ciencia e Cultura* 51:372-384.
- Hodkinson, I. D., and D. Casson. 1991. A lesser predilection for bugs: Hemiptera (Insecta) diversity in tropical rain forests. *Biological Journal of the Linnean Society* 43:101-109.
- Janssens, E. 1953. Contribution à l'étude des Dipteres de Urundi. 4. Asilidae. *Bulletin* 29. Institut Royal des Sciences Naturelles de Belgique, Bruxelles 29:1-15.
- Janssens, E. 1954a. Leptogastrinae (Diptera Asilidae). *Exploration Parc National Upemba Miss. G. F. de Witte, Fascicle* 25:113-134.
- Janssens, E. 1954b. Leptogastrinae (Diptera Asilidae) du Congo Belge. *Annales du Musee Royale du Congo Belge, Sciences Zoologique* 1:400-401.
- Janssens, E. 1957. Contribution à l'étude des leptogastrinae (Diptera Asilidae). *Bulletin* 33(49). Institut Royal des Sciences Naturelles de Belgique, Bruxelles 33:1-12.
- Kress W. J., W. R. Heyer, P. Acevedo, J. Coddington, D. Cole, T. L. Erwin, B. J. Meggers, M. Pogue, R. W. Thorington, R. P. Vari, M. J. Weitzman, and S. H. Weitzman. 1998. Amazonian biodiversity: assessing conservation priorities with taxonomic data. *Biodiversity and Conservation* 7:1577-1578.
- León-Cortés J. L., M. J. Soberón, and J. B. Llorente. 1998. Assessing completeness of Mexican sphinx moth inventories through species accumulation curves. *Diversity and Distributions* 4:37-44.
- Londt, J. G. H. 1985. Afrotropical Asilidae (Diptera) 12: the genus *Neolopbonotus* Engel, 1925. Part 1. The *chiontrix*, *squamosus* and *angustibarbus* species-groups (Asilinae: Asilini). *Annals of the Natal Museum* 27:39-114.
- Londt, J. G. H. 1988. Afrotropical Asilidae (Diptera) 15: the genus *Neolopbonotus* Engel, 1925. Part 4. The *comatus* species-group (Asilinae: Asilini). *Annals of the Natal Museum* 29:1-166.
- Maunder, M., B. Lyte, J. Dransfield, and W. Baker. 2001. The conservation value of botanic garden palm collections. *Biological Conservation* 98:259-271.
- May, R. M. 1988. How many species are there on earth? *Science* 241:1441-1449.
- McAlpine, D. K. 1994. Review of the species of *Achias* (Diptera: Platystomatidae). *Invertebrate Taxonomy* 8:117-281.
- McKamey, S. H. 1999. Biodiversity of tropical Homoptera, with the first data from Africa. *American Entomologist* 45:213-222.
- Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403:853-858.
- Novotný, V., Y. Basset. 2000. Rare species in communities of tropical insect herbivores: pondering the mystery of singletons. *Oikos* 89:564-572.
- Ødegaard, F. 2000. How many species of arthropods? Erwin's estimate revisited. *Biological Journal of the Linnean Society* 71:583-597.
- Papp, L. 1978. Contribution to the revision of the Palaearctic Lauxaniidae (Diptera). *Annales Historico-Naturales Musei Nationalis Hungarici* 70:213-231.
- Petersen, J. F. T., R. Meier, and M. N. Larsen. 2003. Testing species richness estimation methods using museum label data on the Danish Asilidae. *Biodiversity and Conservation* 12:687-701.
- Platnick, N. 1999. Dimensions of biodiversity: targeting megadiverse groups. Pages 33-52 in J. Cracraft and F. T. Grifo, editors. *The living planet in crisis: biodiversity science and policy*. Columbia University Press, New York.
- Ponder, W., and D. Lunney, editors. 1999. *The other 99%: the conservation and biodiversity of invertebrates*. Transactions of the Royal Zoological Society of New South Wales, Mosman, Australia.
- Ponder, W. F., G. A. Carter, P. Flemons, and R. R. Chapman. 2001. Evaluation of museum collection data for use in biodiversity assessment. *Conservation Biology* 15:648-657.
- Rose, A. 2000. The rove beetles of the collection F and R. Struve from Borkum Island of the North Sea. *Entomologische Blätter für Biologie und Systematik der Käfer* 96:127-156.
- Scoble, M. J., K. J. Gaston, and A. Crook. 1995. Using taxonomic data to estimate species richness in Geometridae. *Journal of the Lepidopterists' Society* 49:136-147.
- Smith, D. 2001. Systematic biology initiative. *Antenna*, Bulletin of the Royal Entomological Society 25:268-272.
- Soberon J., and J. Llorente. 1993. The use of species accumulation functions for the prediction of species richness. *Conservation Biology* 7:480-488.
- Soberón, J. M., J. B. Llorente, and L. Oñate. 2000. The use of specimen-label databases for conservation purposes: an example using Mexican papilionid and pierid butterflies. *Biodiversity and Conservation* 9:1441-1466.
- Steyskal, G. C. 1965. Trend curves of the rate of species description in zoology. *Science* 149:880-882.

- Stork, N. E. 1988. Insect diversity: facts, fiction and speculation. *Biological Journal of the Linnean Society* **35**:321-337.
- Stork, N. E. 1993. The biodiversity crisis: an agenda for global research. *Biology International* **1993**:59-64.
- Wheeler, Q. D. 1990. Insect diversity and cladistic constraints. *Annals of the Entomological Society of America* **83**:1031-1047.
- Wheeler, Q. D. 1995. Systematics, the scientific basis for inventories of biodiversity. *Biodiversity and Conservation* **4**:476-489.
- Williams, P. H., C. R. Margules, and D. W. Hilbert 2002. Data requirements and data sources for biodiversity priority area selection. *Journal of Bioscience* **27**(supplement 2):327-338.

#### Appendix 1. *Euscelidia* species in biodiversity hotspots.\*

Western African Forests: *E. artapbernes* (1; 1; 1; no); *E. datis* (2; 2; 2; no); *E. discors* (4; 1; 1; yes); *E. lata* (15; 3; 3; no); *E. milva* (3; 1; 1; no); *E. moyoensis* (13; 1; 2; no); *E. procula* (12; 10; 10; no).

Cape Floral Province: *E. brunnea* (25; 12; 13; no); *E. capensis* (13; 2; 6; yes).

Eastern Arc: *E. artapbernes* (1; 1; 1; no); *E. procula* (19; 7; 12; no); *E. pulchra* (2; 2; 2; no); *E. tsavo* (1; 1; 1; no).

Indo-Burma: *E. lepida* (1; 1; 1; yes); *E. livida* (1; 1; 1; yes); *E. popa* (1; 1; 1; yes).

Madagascar: *E. fastigium* (1; 1; 1; yes).

Mediterranean Basin: *E. pallasii* (55; 5; 27; no).

Philippines: *E. rapacoides* (1; 1; 1; yes).

Western Ghats/Sri Lanka: *E. abbreviata* (1; 1; 1; yes); *E. cobice* (9; 2; 2; yes); *E. flava* (5; 2; 2; yes); *E. glabra* (27; 1; 1; yes); *E. marion* (41; 15; 15; no); *E. prolata* (3; 1; 2; yes); *E. splendida* (19; 1; 1; yes).

\*Information in parentheses refers, respectively, to number of specimens; number of localities; number of collecting events; and endemism of species (detailed specimen records given in Dikow [2003]).

