[1]*Departamento de Zoología, Universidad de Santiago de Compostela, Santiago de Compostela, Spain;* [2]*Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC), Madrid, Spain;* [3]*NERC, Centre for Population Biology, Division of Biology, Imperial College London, Ascot, UK;* [4]*Natural History Museum and Biodiversity Research Center, The University of Kansas, Lawrence, KS, USA*

Assessing *alpha* and *beta* taxonomy in eupelmid wasps: determinants of the probability of describing good species and synonyms

A. Baselga[1], J. M. Lobo[2], J. Hortal[3], A. Jiménez-Valverde[4] and J. F. Gómez[2]

Abstract

We analysed the status of the taxonomical knowledge on Afrotropical and Palaearctic Eupelmidae (Hymenoptera, Chalcidoidea), by assessing the historical accumulation of valid new species and synonyms through time, as well as the factors that have determined the process of description of good species and synonyms in these regions. The species accumulation curves for valid species names show that a significant proportion of the taxa remains to be described in both areas. On the contrary, the historical accumulation of synonymic names seems to reach an asymptote in both areas, although the proportion of synonyms compared with the total number of names is higher in the Palaearctic region. We assessed factors influencing the probability of a species being discovered and the probability of being described several times under different names. Both probabilities seem to be affected by the same factors. Distribution range size and geographical location of taxa affect both the processes of discovering good species and producing synonyms in either region. Models using all these predictors explain (1) 36% and 40% of the variance in the year of description of valid names for the Afrotropical and Palaearctic regions respectively, and (2) 56% and 40% of the variation in the number of synonyms per valid species in these two areas. For the species with available information, trophic specialization is also a significant determinant of these probabilities. However, its effect is also accounted for by geographical range size, which can thus be used as a surrogate of the ecological correlates of discovery and redundant description probabilities. In general, the ecologically generalist, widely distributed and early described taxa living in northern and western areas of both regions have been subject of redundant description more times. Further, taxonomical effort is needed in the south and east of the Palaearctic, as well as in the whole of the Afrotropics, where additional effort in a revision of Eupelmid systematics is needed.

Key words: Eupelmidae – Chalcidoidea – synonymy determinants – Afrotropical – Palaearctic – geographical location – distribution range size – trophic range – year of description – hierarchical partitioning

Introduction

A well established taxonomical knowledge is the necessary foundation of any biodiversity research. Thus, the so-called Linnaean shortfall arises as one of the main problems for conservation biogeography (Brown and Lomolino 1998; Whittaker et al. 2005): our knowledge of most of the extant biodiversity is scarce and patchy (May 1988; Colwell and Coddington 1994; Pimm et al. 1995; Stork 1997). The magnitude of this shortfall is not homogeneous, varying between groups of organisms and regions of the world (Gaston 1993; Brown and Lomolino 1998). Because of this, assessing the current state of biotic inventories is a necessary first step to study the patterns of biodiversity, to detect biases in our knowledge and optimize future taxonomical effort (Baselga et al. 2007; Lobo et al. 2007; Hortal et al. 2007, 2008).

Mayr (1969) defined the two initial stages in the historical development of the taxonomical process as *alpha* taxonomy, which is the discovering and naming of new species; and *beta* taxonomy, or the study of the relationships between the already described taxa, through systematic revisionary work of higher taxa. Such monographical revisions allow detecting the species that were redundantly described under different names (i.e. synonyms). Hence, both the historical accumulation of valid species and synonyms and the relationship between both kinds of descriptions could provide information about the status of the taxonomical knowledge in a given group (Fig. 1). Indeed,

any effective assessment of the state of taxonomical knowledge should take into account the degree of development of both stages: a regional inventory can be considered complete only when all the species have been discovered (*alpha* taxonomy) and taxa have been revised as well so that no synonyms remain undetected (*beta* taxonomy). Therefore, the status of the taxonomical process in a given group can be assessed by studying (1) the growth through time of the cumulative number of valid names (Medellín and Soberón 1999; Cabrero-Sañudo and Lobo 2003; Baselga and Novoa 2006; Hortal et al. 2008), (2) the growth of the cumulative number of invalid names (synonyms) and (3) the temporal trends in the proportion of synonyms (see, e.g. Solow et al. 1995 or Alroy 2002).

Apart from the general lack of completeness, historical surveys also suffer from a number of biases related to species-specific characteristics. As a result, different species have different probabilities of being discovered according, e.g. to distribution range size, location, trophic range, showiness, etc. These effects have been described for a number of groups (e.g. Gaston and Blackburn 1994; Blackburn and Gaston 1995; Gaston et al. 1995a,b; Allsopp 1997; Cabrero-Sañudo and Lobo 2003; De Grave 2003; Collen et al. 2004; Diniz-Filho et al. 2005; Gibbons et al. 2005; Baselga et al. 2007; Jiménez-Valverde and Ortuño 2007; Guil and Cabrero-Sañudo 2007). In contrast with the probability of species discovery, the impact of these biases in the probability of being redundantly described under different names remains unknown, although it is commonly acknowledged that undetected synonyms are one of the major sources of bias in taxonomical knowledge (see, e.g. Gaston and Mound 1993). Here we explore for the first time some species characteristics that have an influence on this probability.

Corresponding author: Jose F. Gómez (jgomez@mncn.csic.es)
Contributing authors: Andrés Baselga (andres.baselga@usc.es), Jorge M. Lobo (mcnj117@mncn.csic.es), Joaquín Hortal (j.hortal@imperial.ac.uk), Alberto Jiménez-Valverder (ajvalv@ku.edu)
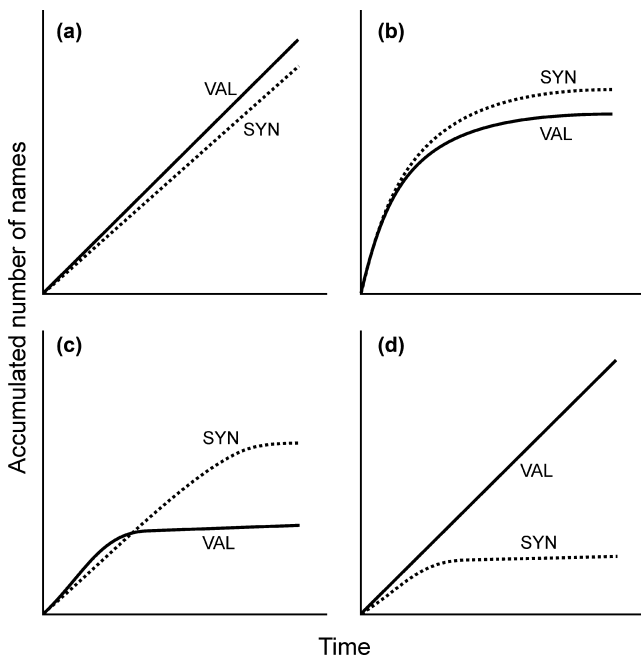
Fig. 1. Conceptual framework showing four hypothetical states in a taxonomical process, inferred from the relative behaviour of the cumulative curves of valid names (VAL) and synonyms (SYN). (a) The number of valid species and synonyms increases with time, implying respectively, that both *alpha* and *beta* taxonomies are still active. (b) Both curves are asymptotic, suggesting that both *alpha* and *beta* taxonomy have been active in the past but not in the present. (c) The discovery of valid species (*alpha* taxonomy) has already finished in the past, whereas synonyms described later were identified and invalided, implying the existence of a strong *beta* taxonomical activity. (d) New species are still being described (*alpha* taxonomy) whereas no synonyms are currently being described, implying either a lack of *beta* taxonomy (i.e. redundant descriptions are still considered valid because of the reduced revisionary work) or an excellent efficiency of *alpha* taxonomy (i.e. all new species are valid and none is redundantly described)

To exemplify these historical trends in taxonomical knowledge we focussed on Eupelmidae, which is one of the largest families of Chalcidoidea (Hymenoptera, 'Parasitica'), with 907 recognized species distributed in all world regions (Gibson 1995; Noyes 2003). The vast majority of eupelmid wasps are ectoparasitoids and facultative hyperparasitoids on the immature stages of other insects, with hosts recorded in Lepidoptera, Homoptera, Hymenoptera, Coleoptera, Neuroptera and Orthoptera (Noyes 2003). In spite of Eupelmidae being one of the insect groups hardest to identify, partly because of their small size, there is a significant amount of taxonomical work already available for this family. After the seminal monograph of Gibson (1995), several papers have been published adding new taxa and information onto the eupelmids, as well as establishing new synonymies (Yang 1996; Sheng et al. 1997; Sheng 1998; Askew and Nieves-Aldrey 2000, 2004, 2006; Delvare 2001; Anitha 2004; Askew 2005).

In this study we analyse the status of taxonomical knowledge on eupelmid wasps in two different biogeographical regions (Palaearctic and Afrotropics), which should *a priori* differ in the level of *alpha* and *beta* taxonomy as mirrored in current catalogues. We considered two biogeographical regions with different status of taxonomical knowledge to discern how much of these effects are general for all

taxonomical processes, and how much can be attributed to the peculiarities of each regional fauna. In sum, our specific aims are: (1) to analyse the degree of completeness and the reliability of the Afrotropical and Palaearctic inventories of Eupelmidae; (2) to describe the temporal trends in description of valid species and synonyms in either region; (3) to assess the relative importance of several characteristics of eupelmid species as determinants of the probability to discover new species and of a species being redundantly described several times under different names. We use these results to identify the regions where future surveys should be located, as well as to forecast the characteristics of currently undetected synonyms.

Data and methods

Taxonomical data

One hundred and fifty-four eupelmidae species have been described from the Afrotropical region and 192 from the Palaearctic, including the more recent systematical revisions (Kalina 1981, 1988; Bouček 1988; Gibson 1995; Askew and Nieves-Aldrey 2000, 2004, 2006; Noyes 2003). The boundaries of the biogeographical regions were defined according to Vigna Taglianti et al. (1992) and Cox (2001). Herein, we refer to the species names currently considered as being good species as *valid names*, and to those names that are now considered invalid as *synonyms*. Taxonomical and world geographical information for all the species currently considered as valid follows Gibson (1995) and Noyes (2003).

Historical accumulation curves of valid names and synonyms

We extracted the year of description of all valid names (herein, VAL), as well as the year of description of the names that are currently considered synonyms (herein, SYN). We also measured the temporal variation in the *taxonomical efficiency* through time as the proportion of the cumulative total number of names described until a given year that are now regarded as synonyms (%SYN). We plotted against the year the accumulated number of valid names and synonyms, as well as %SYN, to examine the temporal trends in the names that become accepted or unaccepted (cf. Medellín and Soberón 1999). The cumulative number of species was plotted against the year and different types of species-accumulation functions were used without success to fit the resultant curve (Flather 1996). Finally, we arbitrarily divided the data into periods of similar description trends, after examining the historical accumulation of valid names and synonyms (see also Hortal et al. 2008).

Temporal trends in alpha and beta taxonomy

We evaluate the variation in the effectiveness of the taxonomical work and the completeness of the checklist of each region by comparing (1) the rate of description of valid species, measured as the accumulation of valid names per time period (herein, RVAL); (2) the rate of description of invalid names, measured as the accumulation of synonyms per time period (herein, RSYN); and (3) taxonomical efficiency, measured as %SYN. We also use %SYN as a measure of the relationship between descriptive (*alpha*) and revisionary (*beta*) taxonomy. Our assumption is that, the more taxonomical revisions are carried out, the higher is the probability for a given species name to be synonymized. Given that the synonyms are allocated to the date when the name was first introduced, rather than the date when it was synonymized, %SYN will unavoidably show a diminishing trend with time, as newly described species will have had less time to be reviewed and eventually synonymized. Irrespective of that, the steepness of the decay of this percentage through time can help us to measure the quality of currently valid names. We used linear regressions to relate RVAL and RSYN to the year, to characterize trends in *alpha* and *beta* taxonomy. We compared the slopes of these regressions ($\pm 95\%$ confidence interval) between the two biogeographical regions, as well as between the arbitrarily defined time intervals.

Determinants of the probability of discovery and redundant description

We explored the effects of several variables (1) on the temporal variations in the probability to describe a new valid species (measured as VAL) in each region, as well as (2) on the probability of redundant description of species (measured as number of synonyms per valid species, NSYN). Five predictors were used to model VAL: distribution range size, maximum longitude, minimum longitude, maximum latitude and minimum latitude. These five variables plus the year of description were used to model NSYN. Year of description was extracted from the published literature and Noyes (2003); maximum and minimum latitude and longitude were estimated using the available distribution data for each species (Gibson 1995; Noyes 2003); a crude measure of area (i.e. distribution range size) was calculated in a GIS as the number of terrestrial geographical degree cells comprised within the bounds provided by the maximum and minimum latitude and longitude (excluding cells with less than 15% land surface).

The relationships between VAL, NSYN and the abovementioned predictors were analysed using General Linear Models (GLM; Legendre and Legendre 1998) in R (R Development Core Team 2006). The possibility of complex relationships between these predictors and both response variables was explored by means of hierarchical partitioning analyses (Chevan and Sutherland 1991; Mac Nally 2002). Here, the explanatory variables identified as significant in the former analyses were explored to identify the portions of the variation in the dependent variable that are accounted for by the independent and shared effects of the predictors. We used the hierarchical partitioning algorithm of Chevan and Sutherland (1991), which considers all possible models (i.e. combinations of variables) in a multiple regression framework to estimate the independent fraction of variability accounted for by each predictor. Such fraction is calculated as the average model fit (measured in our case as R^2) generated by the inclusion of each predictor in all the possible models in which it appears (Mac Nally 2000, 2002; Mac Nally and Horrocks 2002; Quinn and Keough 2002). The result of this analysis is an estimate of the independent (i) contribution of each predictor, as well as its joint (j) contribution with the rest of predictor variables (i.e. the shared variability between two or more predictors that can not be attributed unequivocally to none of them). The statistical significance of the different fractions was calculated by means of the Z score ($i_{observed}$ − $i_{randomized}$/SD$_{irandomized}$); observed scores above the upper 95% level of a standard normal distribution were accepted as significant. Hierarchical partitioning analyses were carried out using the HIER.PART package version 1.0-1 (Walsh and Mac Nally 2005) in R (R Development Core Team 2006).

We also conducted a complementary analysis to assess the ecological correlates of the probabilities of discovery and redundant description. This analysis was independent of the former because here we were limited by the low number of species for which reliable ecological information is available (41 in the Afrotropical and 100 in the Palaearctic region). The trophic specialization (i.e. host specificity) of these species was estimated following data extracted from Noyes (2003). We classified the species into two categories of trophic range: stenophagous (species with one host or several hosts belonging to the same insect order) and polyphagous (species with more than two hosts belonging to different insect orders). We used GLM to assess the effect of trophic range in the probability of discovery and redundant species description. As geographical variables can incorporate the effects of ecological traits we have also carried out a backward stepwise selection procedure to estimate the shared variability accounted for by trophic and geographical variables.

Results

Historical processes of description of good species and synonyms

The Afrotropical Eupelmidae constitute approximately 17% of the world species of the family, including 16 species (10.3%) which extend their ranges out of this region (see Noyes 2003). A total of 198 names were employed between 1784 and 2004 to name the eupelmid species from this region. A total of 154 names are currently considered as valid names, described at a rate of 0.79 species per year (see Table 1 and Fig. 2). The 44 names that have been synonymized were erroneously described at a rate of 0.26 species per year, a significantly lower rate than RVAL (see Table 1). Palaearctic eupelmids include 20.8% of the total number of known species (Noyes 2003), of which 25 (13.2%) have been also reported from other regions. A total of 283 names were given to eupelmid taxa from the Palaearctic region between 1783 and 2004. Of these names, 192 are currently valid names, described at a rate of 0.83 species by year (Table 1 and Fig. 2). The 91 names that have been synonymized were described at a significantly higher rate (0.52 species per year) than in the Afrotropics. Nevertheless, while the rate of description of valid species is similar in both regions, the percentage of all used names that were synonymized is slightly higher (32%) in the Palaearctic region (Table 1).

The mean year of description (± 95% confidence interval) of the names that were later synonymized is 1914 ± 11 for the Afrotropical region and 1899 ± 9 for the Palaearctic. Half of the names synonymized in the Afrotropics were proposed before 1863, and 90% before 1952. Redundant descriptions occurred earlier in the Palaearctic, where half of the synonyms were proposed before 1860, and 90% before 1922. As the vast majority of Afrotropical species was described after 1900 (Fig. 2), we have used the historical process of accumulation of valid and synonymized names in that region to define three time spans that correspond to periods of roughly homogenous taxonomical work (Table 1 and Fig 2): (1) until 1905 the rate of description of valid species (RVAL) is low and the proportion of synonyms (%SYN) is high (65%); (2) from 1906 to 1960 RVAL is highest and %SYN diminishes drastically, as only 18% of the species described during this

	Until 1905	1906–1960	1961–2004	Total
Afrotropical				
% synonyms	65	18	13	22
Rate valid names	0.06 (0.05–0.06)	2.02 (1.80–2.24)	0.62 (0.48–0.76)	0.79 (0.73–0.85)
Rate synonyms	0.12 (0.11–0.13)	0.51 (0.47–0.55)	0.10 (0.08–0.12)	0.26 (0.25–0.27)
Palaearctic				
% synonyms	43	39	4	32
Rate valid names	0.42 (0.41–0.43)	1.35 (1.25–1.45)	1.29 (1.19–1.39)	0.83 (0.81–0.85)
Rate synonyms	0.39 (0.37–0.41)	1.01 (0.94–1.08)	0.05 (0.04–0.06)	0.52 (0.50–0.54)

Table 1. Percentage of synonyms on the total number of names, and rates of description (number of species by year) of valid names and synonyms

Results are shown for the two biogeographical regions analysed in the three periods studied, as well as for the total taxonomical process carried out between 1748 and 2004 (see text). Rates were calculated as the linear regression between the year and the cumulative number of valid names and synonyms respectively, in each region and/or period. The 95% confidence interval of all estimates is given in brackets.
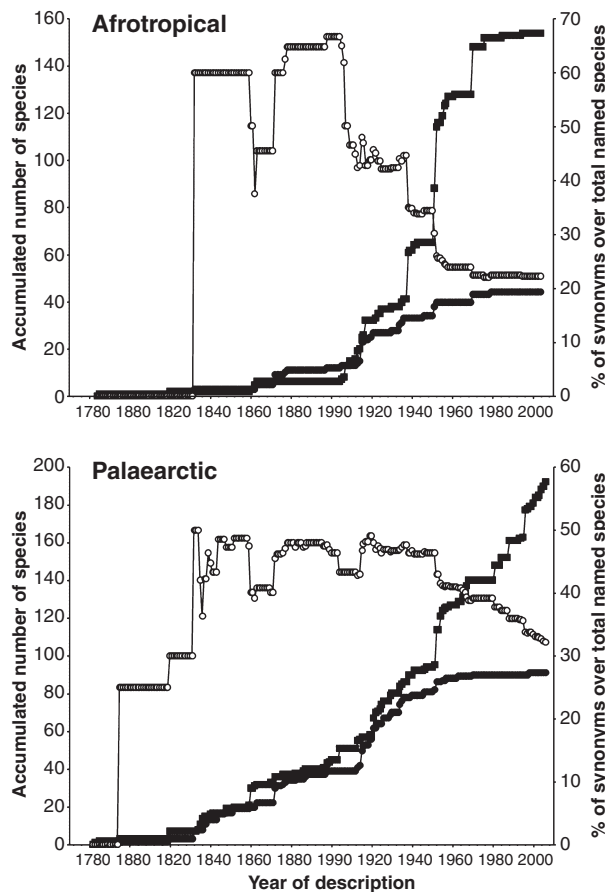
Fig. 2. Historical process of accumulation of valid names (dark squares) and synonyms (dark circles) in the Afrotropical and Palaearctic Eupelmidae, according to the date of their description. The percentage of synonyms over the total number of names (%SYN) is also shown (open circles)

period have been synonymized; (3) from 1961 to present day (2004) both rates decay to their lowest levels. In the Palaearctic region the taxonomical process follows a similar general temporal trend, although with a higher rate of description of valid names both in recent and in past times, and a higher rate of description of synonyms (RSYN) until 1905 (see Fig. 2 and Table 1). No asymptotic function could be satisfactorily adjusted to the cumulative number of valid names (not shown), suggesting that knowledge on eupelmid diversity is far from being complete in both regions.

Determinants of the probability of discovery and redundant description

Most of the analysed predictors were significantly related to the year of first description of species under their valid names (Table 2, Fig. 3). The larger the species distribution range, and the higher its maximum latitude and the lower its minimum longitude (i.e. the more western the species is located), the earlier the species was described in both biogeographical regions. In the Palaearctic region minimum latitude was positively related to VAL (later descriptions at southern localities), whereas in the Afrotropics the maximum longitude was negatively related to VAL (earlier descriptions at westernmost localities). The partition of VAL variation

among the predictors (Table 3) showed that most of the variability accounted for in the Afrotropics is contributed by several variables, being all independent fractions lower than 10%, except for distribution range size, which shows a remarkable independent effect (15%). In the Palaearctic, the results are similar, although in this case maximum latitude is the most remarkable predictor, explaining nearly a fifth of total variability (20%). Models including all significant variables accounted for 56% of total variability in VAL for Afrotropical species, 40% in the case of the Palaearctic (Table 2).

The majority of the predictors analysed were also significantly related to the number of synonyms per valid species (Table 4, Fig. 4). Here, the larger the size of the distribution range of a species, the earlier it was described, and the higher its maximum latitude and longitude and the lower its minimum longitude, the higher is NSYN in both biogeographical areas. Minimum latitude is also negatively related to the number of synonyms in the Palaearctic. Most variation in NSYN explained by any of the predictors in the Afrotropics is shared with other variables (Table 3), the most relevant independent effect being that of distribution range size (22%). In the Palaearctic, both this variable and the year of description stand out as important independent predictors of NSYN (Table 3). Models including all significant variables were able to explain 56% of the variability of NSYN in the Afrotropical region, 40% in the Palaearctic (Table 4).

The assessment of the ecological correlates of the probabilities of discovery and redundant description showed that polyphagous species were described significantly earlier than stenophagous ones both in the Afrotropical ($F_{1, 39} = 11.10$, $p = 0.0019$) and the Palaearctic region ($F_{1, 98} = 16.40$, $p = 0.0001$). Polyphagous species also have a significantly higher number of synonyms than stenophagous ones ($F_{1, 39} = 5.52$, $p = 0.0239$ for the Afrotropical and $F_{1, 98} = 12.97$, $p = 0.0005$ for the Palaearctic region). Moreover, the variation in the year of description and number of synonyms explained by trophic range is also accounted for by distribution range size. In both regions, models with

Table 2. Relationship between several predictor variables and the year of description of valid names (VAL) in the two biogeographical regions studied

Variable	%Expl.Var.	F	d.f.
Afrotropical species			
Distribution range size	33.8	77.49*	1, 152
Maximum longitude	15.4	27.62*	1, 152
Minimum longitude	13.9	24.47*	1, 152
Maximum latitude	21.8	42.31*	1, 152
Minimum latitude	0.002	0.37	1, 152
Whole Model (all predictors)	35.6	20.56*	4, 149
Palaearctic species			
Distribution range size	20.6	49.37*	1, 190
Maximum longitude	0.01	1.96	1, 190
Minimum longitude	24.6	61.84*	1, 190
Maximum latitude	34.1	98.09*	1, 190
Minimum latitude	7.4	15.25*	1, 190
Whole model (all predictors)	40.0	31.15*	4, 187

%Expl.Var. is the percentage of explained variance (R^2); F is the F statistic (*means significant at a $p < 0.0001$ level) and d.f. is the number of degrees of freedom. All relationships identified are linear.
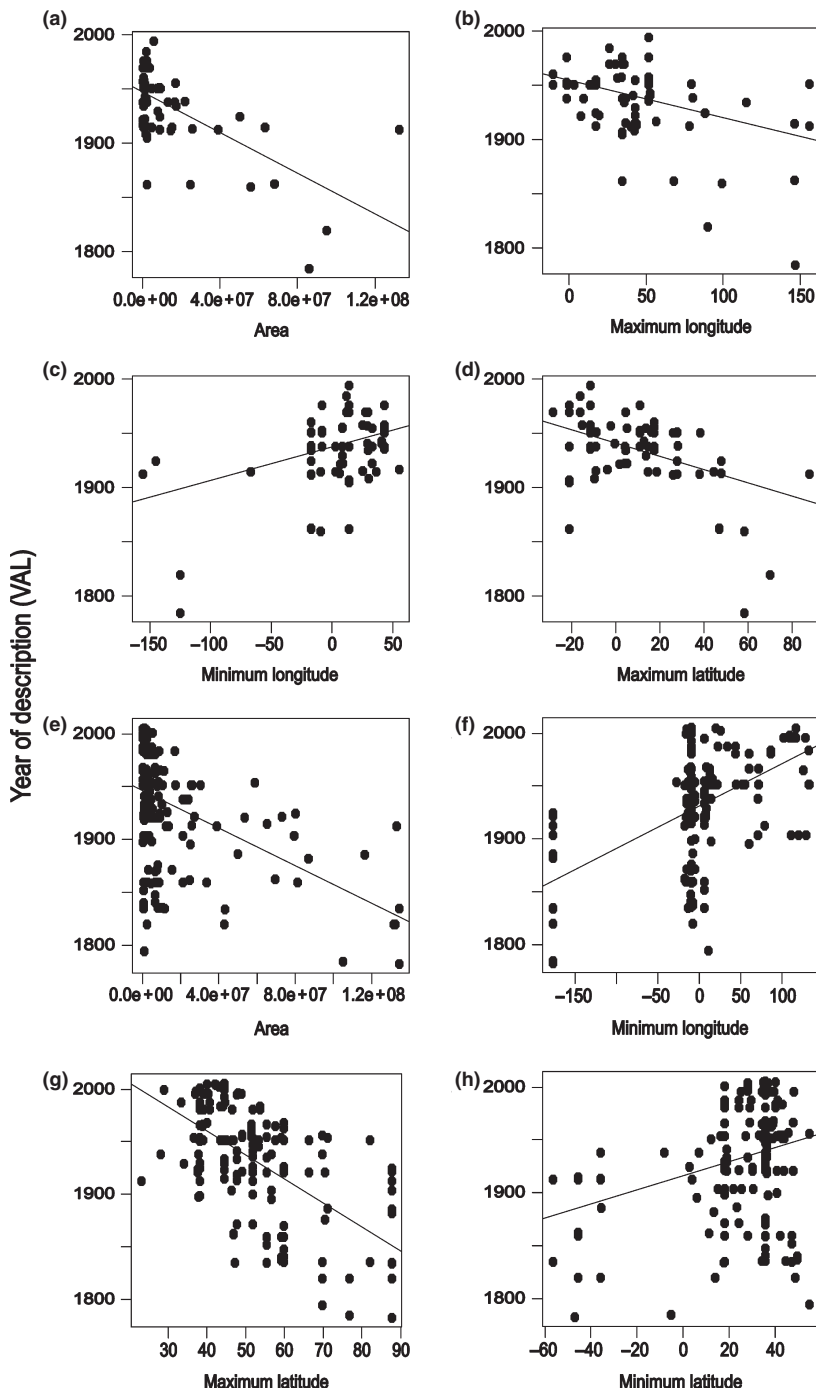
Fig. 3. Relationship between the year of description of valid species names and the statistically significant predictors (see Table 2) in the Afrotropical (a–d) and Palaearctic (e–h) regions

distribution range size and trophic range do not explain significantly more variability than models with distribution range size only; the two variables explain 36% of variation in the year of description for the Afrotropical region and 22% for the Palaearctic (compared to 33% and 21% respectively, explained by distribution range size alone); and 54% of variation in the number of synonyms for the Afrotropical and 35% for the Palaearctic region (compared with 53% and 33% respectively, explained by distribution range size alone). Thus, trophic range drops from all the models when a backward stepwise procedure is used with these two variables, and it can be assumed that the independent contribution of this variable is negligible.

Discussion

Our results point out that the Eupelmidae inventory is far from being complete in both the Afrotropical and Palaearctic regions. Although the cumulative curve of valid names from the Afrotropics seems to reach an asymptotic shape in its last time interval, this observed trend is likely caused by sampling and taxonomical deficiencies. The accumulation of taxonomical work carried out in this region is based on monographs concentrated in short periods, separated by long intervals without active work on eupelmid systematics. Thus, the apparently asymptotic trend in the rate of species description is more likely to be an artefact than a sign of the completeness

Table 3. Hierarchical partitioning of the determinants of valid and redundant description for both regions

Variable	i	j	Z
Afrotropical species (VAL)			
Distribution range size	15.2	18.6	8.97
Maximum longitude	7.0	8.4	7.68
Minimum longitude	4.9	9.0	4.77
Maximum latitude	8.5	13.2	8.52
Palaearctic species (VAL)			
Distribution range size	7.5	13.1	9.34
Minimum longitude	9.2	15.4	10.26
Maximum latitude	19.6	14.4	24.72
Minimum latitude	3.6	3.8	4.24
Afrotropical species (NSYN)			
Year of description	6.5	16.5	4.99
Distribution range size	21.6	32.4	6.79
Maximum longitude	10.4	14.2	4.69
Minimum longitude	10.3	18.4	9.55
Maximum latitude	7.1	18.5	7.37
Palaearctic species (NSYN)			
Year of description	10.2	14.4	10.38
Distribution range size	10.6	22.0	21.87
Maximum longitude	2.5	5.0	2.08
Minimum longitude	5.1	14.0	6.59
Maximum latitude	5.7	14.9	7.90
Minimum latitude	6.0	13.3	9.19

Percentage of variability accounted for by independent (i) and joint (j) contributions of the predictors formerly identified as significant predictors of the year of description of valid names (VAL) and the number of synonyms per valid species (NSYN). Z is the Z score used to assess significance of the independent effects; all terms were significant at a $p < 0.02$ level.

Table 4. Relationships between several predictor variables and the number of synonyms per valid species (NSYN) in the two biogeographical regions studied

Variable	% Expl.Var.	F	d.f.
Afrotropical species			
Year of description	23.0	45.31*	1, 152
Distribution range size	54.0	178.4*	1, 152
Maximum longitude	24.6	49.65*	1, 152
Minimum longitude	28.7	61.18*	1, 152
Maximum latitude	25.6	52.32*	1, 152
Minimum latitude	0.1	0.202	1, 152
Whole Model (all predictors)	55.9	37.50*	5, 148
Palaearctic species			
Year of description	24.5	61.62*	1, 190
Distribution range size	32.5	91.63*	1, 190
Maximum longitude	7.5	15.42*	1, 190
Minimum longitude	19.2	44.99*	1, 190
Maximum latitude	20.6	49.33*	1, 190
Minimum latitude	19.4	45.62*	1, 190
Whole Model (all predictors)	40.1	20.67*	6, 185

%Expl.Var. is the percentage of explained variance (R^2); F is the F statistic (*significant at a $p < 0.001$ level) and d.f. is the number of degrees of freedom. All relationships identified are linear.

of the inventory. On the contrary, the curve of valid species from the Palaearctic is not asymptotic because of the high number of species that have been described and/or recorded in recent years. In spite of the higher level of taxonomical knowledge for this region, it is evident that the process of inventory is not yet complete here either.

Regarding the historical evolution of *alpha* taxonomy in either region, three clear periods can be distinguished in the Afrotropics, where the rate of description of valid species was very low until the early 20th century, high between 1905 and 1960, and low again from 1960 to present (2004) (Fig. 2). On the contrary, *alpha* taxonomical activity is quite uniform across time in the Palaearctic region (Fig. 2). This difference may be related to the intensive sampling effort in some areas of Europe because of the proximity of recorders' home ranges (Dennis and Thomas 2000) and the general lack of local taxonomists in the Afrotropics after the decolonization of African countries.

The historical trends in the rate of description of synonyms inform at the same time about the effectiveness of *alpha* taxonomical work and the existence of *beta* taxonomical activity. This is because the number of synonyms in a given year depends both on the errors of *alpha* taxonomy in that year, and on the subsequent identification of this errors through revisionary work (*beta* taxonomy). For this reason, the stabilization of the number of synonyms in either region during the later period could be interpreted both as an increment in the effectiveness of *alpha* taxonomy or as the lack of *beta* taxonomical activity (see Fig. 1d). In the Afrotropics the proportion of synonyms decayed much earlier (about 1900) compared with the Palaearctic region (about 1950). If the effectiveness of *alpha* taxonomy was similar in both regions within the same historical period, such difference between trends in the proportion of synonyms Palaearctic and Afrotropical regions would suggest that the early decay in the Afrotropics is mostly because of the lack of *beta* taxonomical effort.

Our results can also help identifying the characteristics of the eupelmid species yet to be discovered (see e.g. Gaston et al. 1995a,b; Allsopp 1997; Cabrero-Sañudo and Lobo 2003; Collen et al. 2004; Gibbons et al. 2005; Baselga et al. 2007). Among the geographical factors, the size of the distribution range stands out as a determinant of the probability of description. Widely distributed Eupelmidae species were described before than restricted ones. This factor appears recurrently as a significant predictor of the rates of species discovery of many groups (Gaston et al. 1995a,b; Allsopp 1997; Cabrero-Sañudo and Lobo 2003; Collen et al. 2004; Gibbons et al. 2005; Baselga et al. 2007; Jiménez-Valverde and Ortuño 2007), evidencing the apparently naïve fact that widespread and common species are the easiest to inventory, hence they have been the first being discovered.

Apart from range size, the probability of description presents some consistent geographical patterns, probably caused by both the geographically biased allocation of the available taxonomical resources, and the unequal distribution of species. Here, we argue that (1) the historical distribution of taxonomists, traditionally concentrated in northwestern temperate countries, and to a lesser extent (2) the asymmetries in the spatial location of the biomes in both regions, are responsible for the geographical patterns in the historical, taxonomical process evidenced by our results. The concentration of taxonomists in Northern Europe implied that the species reaching northernmost regions were described first in the Palaearctic region (Fig. 3g) but also in the Afrotropics (Fig. 3d). In the latter, the species living in Africa but also present in Europe were described first, and exclusively Afrotropical species reaching latitudes closer to Europe were described earlier than those living more south. Regarding the spatial disposition of biomes, tropical biomes placed out of the Afrotropical region are situated both eastwards (SE Asia,
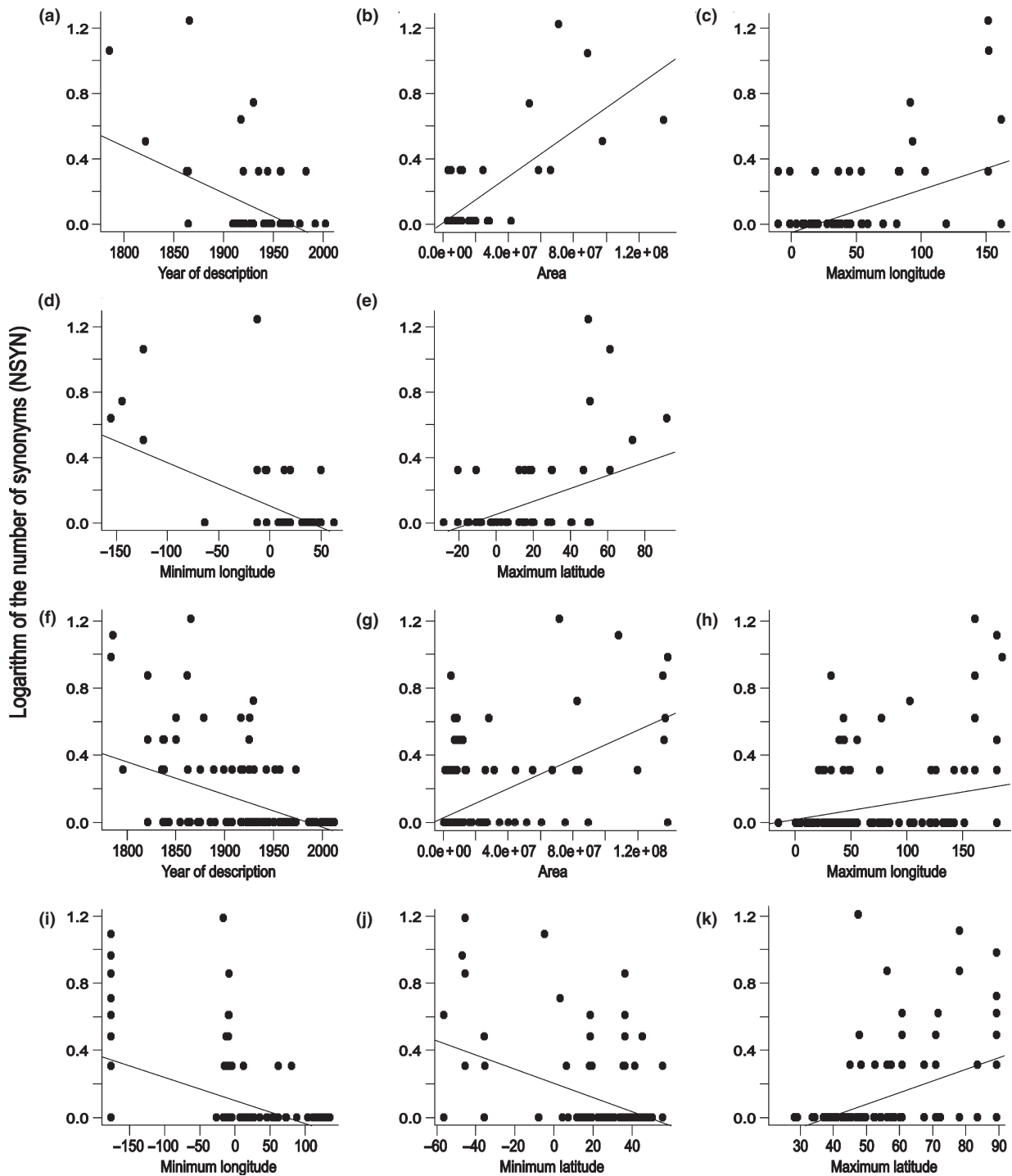
Fig. 4. Relationship between the number of synonymized names per valid name and the statistically significant predictors (see Table 3) in the Afrotropical (a–e) and Palaearctic (f–k) regions

Australia) and westwards (South America), whereas the Palaearctic has only a westwards temperate counterpart (North America). For this reason, minimum longitude is a relevant predictor in both regions (Fig. 3c, f), because Afrotropical species also living in South America and Palaearctic species also living in North America were described first, pinpointing the influence of American taxonomists on the European systematics of this group. In contrast, maximum longitude was an important predictor of the year of description in the Afrotropics (Fig. 3b) but not in the Palaearctic, also

because of the relative position of land masses. Afrotropical species reaching SE Asia, Australia or the islands of the Indian Ocean were described first independently of the size of their ranges, because these regions have received more taxonomical effort than the African continent. Another asymmetry derives from the fact that the Palaearctic region has a temperate counterpart in the Southern hemisphere, whereas tropical biomes are concentrated near the equator. Thus, minimum latitude is only a significant predictor in the Palaearctic region, because species living in the Palaearctic but also in the

temperate regions of the southern hemisphere were described first. Finally, although the effect of range size on the probability of discovery is fairly similar in both continents, the independent relevance of the geographical location (maximum latitude especially) is higher in the Palaearctic. This result suggests that the influence of the proximity of the species to the taxonomist's home site in the description of valid species names is higher in the Palaearctic; species living in areas with higher concentrations of taxonomists are described earlier independently of range size.

Perhaps the most enlightening of our results is that geographical location and distribution range size of the species are related not only to their probability of being described, but also to the probability of being redundantly described. In other words, the same factors that increase the chance of a species of being discovered by a given taxonomist also increase its chance of being independently 'discovered' by another taxonomist. This fact points out that eupelmid taxonomy is still in a primary stage in which *alpha* taxonomy is much more active than *beta* taxonomy: the lack of general revisions is the main factor preventing taxonomists from recognizing already described species, thus increasing the probability of repeatedly describing a species under different names. In both regions, the distribution range size appears as the main determinant of the probability of a species being redundantly described. Finally, the most important difference between the two regions is the independent effect of the year of description on the chance of having received synonyms. Since the year of description covariates with the range size and geographical location, the shared fractions could be attributed to any of these variables. However, the independent effect of the year of description would be indicating the unavoidable role of time in increasing the probability of receiving synonyms (i.e. species recently described could not have had the chance of receiving synonyms yet), regardless of any other characteristic of the species. However, to generate this purely temporal trend, *beta* taxonomy should have been active, allocating synonyms to their senior names. Thus, the larger independent fraction of the variation in the number of synonyms explained by the year of description in the Palaearctic region suggests that *beta* taxonomical processes have been more effective in the Palaearctic than in the Afrotropics. Combining this conclusion with the higher proportion of synonyms in the Palaearctic region (Table 1) and the early stabilization of the synonyms' curve in the Afrotropics, it seems clear that further *beta* taxonomical work is still needed in the latter region.

Our assessment of the ecological (i.e. trophic) correlates confirms the relevance of taxonomical biases from an ecological or evolutionary point of view. Our results show that generalist species have higher probabilities to be discovered and to become the subject of redundant descriptions than specialist species in either region, and, importantly, that simple geographical variables (as distribution range size) are good surrogates of these ecological characteristics. Therefore, in any taxonomical group with deficient taxonomy (which is the rule for most invertebrate taxa) an *a priori* hypothesis should be made that any inferred evolutionary or ecological pattern could be biased. Hence, any large-scale analysis should first rule out the possibility that the studied ecological traits have determined the probability of discovery and/or redundant description of the species, or at least quantify the extent of these biases.

In the light of these results, it is evident that an important effort on *alpha* and *beta* taxonomy is yet needed to improve the quality of the current taxonomical knowledge on the eupelmids in particular, and hyperdiverse and poorly known groups in general. In the case of *alpha* taxonomy, the lack of asymptotic trends in the cumulative curves of valid names shows that a number of species remain to be discovered, even in the 'well-known' Palaearctic region. In the case of *beta* taxonomy, it remains clear that the results of *alpha* taxonomical studies should be evaluated in the future, and that species characteristics influence the probability of being redundantly described. As synonyms appear mainly within the species described in early times, ecologically and morphologically similar species that were described long time ago in different regions and/or by different taxonomists should be prioritized for the reassessment of their taxonomical status. Undetected synonymy could have severe implications in conservation biogeography and biodiversity studies, as estimations on the number of species of some taxa can be artificially inflated (Solow et al. 1995; May 1990), and the measures of rarity and species replacement can be dramatically flawed. Therefore, additional effort in *alpha* and *beta* taxonomy should be encouraged and funded.

Acknowledgments

Resumen

Evaluación de la taxonomía alfa y beta en eupélmidos: determinantes de la probabilidad de descripción de buenas especies y sinónimos

Se analiza el estado del conocimiento taxonómico sobre los Eupelmidae (Hymenoptera, Chalcidoidea) afrotropicales y paleárticos, mediante la evaluación de la acumulación histórica de nuevas especies válidas y sinónimos (descripciones redundantes de taxones previamente conocidos), así como los factores que han determinado el proceso de descripción de buenas especies y sinónimos en ambas regiones. Las curvas históricas de acumulación de especies válidas muestran que una proporción significativa de especies permanece sin describir en ambas áreas. Por el contrario, la acumulación histórica de sinónimos parece haber alcanzado una asíntota en las dos regiones, aunque la proporción de sinónimos sobre el total de nombres utilizados es mayor en la región paleártica. Se analizan también los factores que influyen en la probabilidad de que una especie sea descubierta y en la probabilidad de que sea redundantemente descrita varias veces, produciendo sinónimos. Ambas probabilidades parecen estar afectadas por los mismos factores. El tamaño del área de distribución y la localización geográfica de las taxones afecta tanto al proceso de descubrimiento de especies válidas como a la producción de sinónimos en ambas regiones. Los modelos construidos usando las citadas variables predictoras explican (i) el 36 y 40% de la varianza en el año de descripción de nombres válidos en las regiones afrotropical y paleártica, respectivamente, y (ii) el 56 y 40% de la varianza en el número de sinónimos por especie válida en esas dos regiones. Para las especies con información disponible, el grado de especialización trófica es también un determinante significativo de dichas probabilidades. Sin embargo, el tamaño del rango de distribución da cuenta del efecto de estas variables, por lo que puede ser utilizado como un indicador de los determinantes ecológicos de las tasas de descripción. En general, las especies generalistas, de amplia distribución y descritas más antiguamente que viven en las regiones más septentrionales y occidentales de ambas regiones han sido objeto de descripciones redundantes más numerosas (por tanto,

poseen más sinónimos). Es necesario un mayor esfuerzo taxonómico en el Sur y Este de la región paleártica, así como en toda la región afrotropical, donde se necesita además un esfuerzo adicional en la revisión sistemática de los Eupelmidae.

References

Allsopp PG (1997) Probability of describing an Australian scarab beetle: influence of body size and distribution. J Biogeogr 24:717–724.

Alroy J (2002) How many named species are valid? Proc Natl Acad Sci USA 99:3706–3711.

Anitha PV (2004) Checklist of Eupelmidae (Hymenoptera: Chalcidoidea) from the Indian subcontinent. In: Rajmohana K, Sudheer K, Girish Kumar P, Santhosh S (eds), Perspectives on Biosystematics and Biodiversity. Prof. T.C. Narendran Commemoration Volume. Systematic Entomology Research Scholars Association, Kerala, pp 617–625.

Askew RR (2005) A new species of Anastatus Motschulsky (Hym., Eupelmidae). Ent Mon Mag 141:215–218.

Askew RR, Nieves-Aldrey JL (2000) The genus Eupelmus Dalman, 1820 (Hymenoptera, Chalcidoidea, Eupelmidae) in peninsular Spain and the Canary Islands, with taxonomic notes and descriptions of new species. Graellsia 56:49–61.

Askew RR, Nieves-Aldrey JL (2004) Further observations on Eupelminae (Hymenoptera, Chalcidoidea, Eupelmidae) in the Iberian Peninsula and Canary Islands, including descriptions of new species. Graellsia 60:27–39.

Askew RR, Nieves-Aldrey JL (2006) Calosotinae and Neanastatinae in the Iberian Peninsula and Canary Islands, with descriptions of new species and a supplementary note on Brasema Cameron, 1884 (Hymenoptera, Chalcidoidea, Eupelmidae). Graellsia 62:87–100.

Baselga A, Novoa F (2006) Diversity of Chrysomelidae (Coleoptera) in Galicia, Northwest Spain: estimating the completeness of the regional inventory. Biodivers Conserv 15:205–230.

Baselga A, Hortal J, Jiménez-Valverde A, Gómez JF, Lobo JM (2007) Which leaf beetles have not yet been described? Determinants of the description of Western Palaearctic Aphthona species (Coleoptera: Chrysomelidae). Biodivers Conserv 16:1409–1421.

Blackburn TM, Gaston KJ (1995) What determines the probabilitiy of discovering a species?: a study of South American oscine passerine birds. J Biogeogr 22:7–14.

Bouček Z (1988) Australasian Chalcidoidea (Hymenoptera). C.A.B. International, Wallingford, UK, pp 832.

Brown JH, Lomolino MV (1998) Biogeography, 2nd edn. Sinauer Associates, Inc, Sunderland, Massachussets, pp 692.

Cabrero-Sañudo FJ, Lobo JM (2003) Estimating the number of species not yet described and their charactetristics: the case of Western Palaearctic dung beetle species (Coleoptera, Scarabaeoidea). Biodiver Conserv 12:147–166.

Chevan A, Sutherland M (1991) Hierarchical partitioning. Am Stat 45:90–96.

Collen B, Purvis A, Gittleman JL (2004) Biological correlates of description date in carnivores and primates. Global Ecol Biogeogr 13:459–467.

Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. Phil Trans R Soc B 345:101–118.

Cox CB (2001) The biogeographic regions reconsidered. J Biogeogr 28:511–523.

De Grave S (2003) Historical patterns in the description of north east Atlantic Decapoda. J Crust Biol 23:669–677.

Delvare G (2001) A revision of genus Pentacladia Westwood (Hymenoptera, Eupelmidae). Rev Fr Ent 23:47–62.

Dennis RLH, Thomas CD (2000) Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. J Insect Conserv 4:73–77.

Diniz-Filho JAF, Bastos RP, Rangel TFLVB, Bini LM, Carvalho P, Silva R (2005) Macroecological correlates and spatial patterns of anuran description dates in the Brazilian Cerrado. Global Ecol Biogeogr 14:469–477.

Flather CH (1996) Fitting species-accumulation functions and assessing regional land use impacts on avian diversity. J Biogeogr 23:155–168.

Gaston KJ (1993) Spatial patterns in the description and richness of the Hymenoptera. In: LaSalle J, Gauld ID (eds), Hymenoptera and Biodiversity. CAB International, Wallingford, UK, pp 277–293.

Gaston KJ, Blackburn TM (1994) Are newly described bird species small-bodied? Biodivers Lett 2:16–20.

Gaston KJ, Mound LA (1993) Taxonomy hypothesis testing and the biodiversity crisis. Proc R Soc B 251:139–142.

Gaston KJ, Blackburn TM, Loder N (1995a) Which species are decribed first?: the case of North American butterflies. Biodivers Conserv 4:119–127.

Gaston KJ, Scoble MJ, Crook A (1995b) Patterns in species description: a case study using Geometridae (Lepidoptera). Biol J Linn Soc 55:225–237.

Gibbons MJ, Richardson AJ, Angel MV, Buecher E, Esnal G, Fernández Alamo MA, Gibson R, Itoh H, Pugh P, Boettger-Schnack R, Thuesen E (2005) What determines the likelihood of species discovery in marine holozooplankton: is size, range or depth important? Oikos 109:567–576.

Gibson GA (1995) Parasitic Wasps of the Subfamily Eupelminae: Classification and Revision of World Genera (Hymenoptera: Chalcidoidea: Eupelmidae). Memoirs on Entomology, International, 5, Gainesville, Florida, USA, pp 421.

Guil N, Cabrero-Sañudo FJ (2007) Analysis of the species description process for a little known invertebrate group: the limnoterrestrial tardigrades (Bilateria, Tardigrada). Biodivers Conserv 16:1063–1086.

Hortal J, Lobo JM, Jiménez-Valverde A (2007) Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife (Canary Islands). Conserv Biol 21:853–863.

Hortal J, Jiménez-Valverde A, Gómez JF, Lobo JM, Baselga A (2008) Historical bias in biodiversity inventories affects the observed realized niche of the species. Oikos 117:847–858.

Jiménez-Valverde A, Ortuño VM (2007) The historical description process of Iberian endemic ground-beetles (Coleoptera, Carabidae): which species are described first? Acta Oecol 31:13–31.

Kalina V (1981) The Palaearctic species of the genus Macroneura Walker, 1837 (Hymenoptera, Chalcidoidea, Eupelmidae), with descriptions of new species. Sb Vedeck Lesnick Ust Vysoke Skoly Zemed v Praze 24:83–111.

Kalina V (1988) Descriptions of new Palaearctic species of the genus Eupelmus Dalman with a key to species (Hymenoptera, Chalcidoidea, Eupelmidae). Silvaec Trop Subtrop 12:3–33.

Legendre P, Legendre L (1998) Numerical Ecology, 2nd edn. Elsevier, Amsterdam, pp 853.

Lobo JM, Baselga A, Hortal J, Jiménez-Valverde A, Gómez JF (2007) How does knowledge of the spatial distribution of species accumulate? Divers Distrib 13:772–780.

Mac Nally R (2000) Regression and model-building in conservation biology, biogeography and ecology: the distinction between – and reconciliation of – 'predictive' and 'explanatory' models. Biodivers Conserv 9:655–671.

Mac Nally R (2002) Multiple regression and inference in ecology and conservation biology: further comments on identifying important predictor variables. Biodivers Conserv 11:1397–1401.

Mac Nally R, Horrocks G (2002) Relative influences of match, landscape and historical factors on birds in an Australian fragmented landscape. J Biogeogr 29:395–410.

May RM (1988) How many species are there on earth? Science 241:1441–1449.

May RM (1990) How many species? Phil Trans R Soc B 330:171–182.

Mayr E (1969) Principles of Systematic Zoology. McGraw-Hill, New York, pp 428.

Medellín RA, Soberón JM (1999) Predictions of mammal diversity of four land masses. Conserv Biol 13:143–149.

Noyes JS (2003) Universal Chalcidoidea Database. World Wide Web electronic publication. http://www.nhm.ac.uk/entomology/chalcidoids [last accessed 10 September 2008].

Pimm SL, Russell GJ, Gittleman JL, Brooks TM (1995) The future of biodiversity. Science 269:347–350.

Quinn GP, Keough MJ (2002) Experimental Design and Data Analysis for Biologists. Cambridge University Press, Cambridge, UK, pp 520.

R Development Core Team (2006) R: a language and environment for statistical computing. Available at http://www.r-project.org, Computer program, version 2.4.1 By R Development Foundation for Statistical Computing, Vienna, Austria [last accessed 15 September 2008].

Sheng JK (1998) A new species of genus *Mesocomys* (Hymenoptera: Eupelmidae) from China. Entomol Sin **5:**26–28.

Sheng JK, Wang GH, Yu YX, Yu JC (1997) Four new species of *Anastatus* Motschulsky (Hymenoptera: Euplemidae) from China. Kun Chong Fen Lei Xue Bao **19:**58–64.

Solow AR, Mound LA, Gaston KF (1995) Estimating the rate of synonymy. Syst Biol **44:**93–96.

Stork NE (1997) Measuring global biodiversity and its decline. In: Reaka-Kudla ML, Wilson DE, Wilson EO (eds), Biodiversity II: Understanding and Protecting our biological Resources. Joseph Henry Press, Washington DC, pp 41–68.

Vigna Taglianti A, Audisio PA, Belfiore C, Biondi M, Bologna MA, Carpaneto GM, de Biase A, de Felici S, Piattella E, Racheli T, Zapparoli M, Zoia S (1992) Riflessioni di gruppo sui corotipi fondamentali della fauna W-paleartica ed in particolare italiana. Biogeographia **15:**159–179.

Walsh C, Mac Nally R (2005) The hier.part Package. R package version 1.01. Available at http://www.cran.r-project.org/web/packages/hier.part/index.html [last accessed 15 September 2008].

Whittaker RJ, Araújo MB, Jepson P, Ladle RJ, Watson JEM, Willis KJ (2005) Conservation Biogeography: assessment and prospect. Divers Distrib **11:**3–23.

Yang ZQ (1996) Parasitic Wasps on Bark Beetles in China (Hymenoptera). Science Press, Beijing, China, pp 336.