



Optimising taxonomic effort to overcome the Linnean shortfall: the case of European leaf beetles

ANDREA FREIJEIRO, CAROLA GÓMEZ-RODRÍGUEZ, ADRIÁN CASTRO-INSUA and ANDRÉS BASELGA Departamento de Zoología, Genética y Antropología Física, Universidad de Santiago de Compostela, Santiago de Compostela, Spain

Abstract. 1. Assessing the factors that govern the probability of a species (i) being discovered and (ii) being described under different names can help to improve the efficiency of future taxonomic efforts.

2. Here, we assess whether species body size, extent of distribution and geographical position affect the probabilities of discovery and redundant description of European Chrysomelidae (Coleoptera).

3. The probability of description of European Chrysomelidae is determined mostly by the characteristics of the species distribution but not by body size; species with broad and northern distributions have been described before.

4. The probability of redundant description is mainly related to the year of description; species with a higher number of synonyms were described earlier and have broader distributions.

5. The extrapolation of the observed trends suggests that undescribed species of Chrysomelidae have narrow distributions and live in southern Europe and that recently described species have a low number of synonyms because the efficiency of leaf beetle taxonomy has increased over time.

Key words. Chrysomelidae, Coleoptera, probability of description, synonyms, taxonomy.

Introduction

Biodiversity faces strong pressure and high species extinction rates similar to those observed during major extinction events in the past (Pimm *et al.*, 1995). While a precise knowledge on global biodiversity is fundamental for conservation, only a fraction of the biodiversity of the planet is known. This lack of knowledge has been termed as the Linnaean shortfall (Brown & Lomolino, 1998; Whittaker *et al.*, 2005; Brito, 2010), and its magnitude is unknown (Hortal *et al.*, 2015). Although estimates of global diversity vary widely, conservative estimates suggest that the total number of species is approximately 5–10 million (e.g. Odegaard, 2000), although only approximately 1.6 million have been named (May, 1988; Hammond, 1992). Curiously, even for mammals, one of the most studied groups, almost 5000

species have been described, but some studies estimate the total number of species to be approximately 8000 (Morell, 1996; Medellín & Soberón, 1999). This scenario is even more precarious for other groups, such as invertebrates (Stork, 1997), which comprise at least 90% of the total number of species on Earth. For arthropods, two third of all arthropod species remain to be discovered and described (May, 2010). Therefore, the Linnaean shortfall is a serious challenge for insects, since most species have not been described yet (Diniz-Filho *et al.*, 2010). Taxonomists are tasked with finding and describing new species. Taxonomy, that is, species discovery and morphology-based description, however, is a time-consuming, highly demanding process. The efficiency of taxonomy depends on several factors such as the selection of sampling areas, the need for well-trained professionals (Gaston & May, 1992; Cracraft, 1996), the amount of resources assigned to systematics (Gaston & May, 1992; Cotterill, 1995) and the taxon-specific factors that govern the probability of discovery (Patterson, 1994; Cabrero-Sañudo & Lobo, 2003). Recent advances provided by molecular techniques (e.g. Hebert

Correspondence: Andrea Freijeiro, Departamento de Zoología, Genética y Antropología Física, Universidad de Santiago de Compostela, c/Lope Gómez de Marzoa s/n, 15782 Santiago de Compostela, Spain. E-mail: andrea.freijeiro@usc.es

et al., 2003; Pons *et al.*, 2006; Gómez-Rodríguez *et al.*, 2015) do not seem to resolve these issues. Due to the aforementioned impediments to taxonomic research, already described species are a non-random subset (Gaston *et al.*, 1995a,b), so our current perspective of biodiversity is quite biased. In addition to the existence of 'unknown' species, the generation of synonyms (i.e. species redundantly described under different names) also increases the magnitude of the Linnean shortfall. In some studies, 20% or more species names within a taxonomic group have been shown to be synonyms (Gaston & Mound, 1993). For example, a ratio of more than 11 names per species has been observed in bumblebees (Williams, 1998).

Multiple reasons can explain the biases and gaps in taxonomic knowledge. Spatial biases arise because the intensity of taxonomic effort is uneven across the world. In some areas, such as central and northern Europe, a long naturalist tradition has ensured that most of their species have been already described. Temporal biases occur when surveys are intensely conducted, but lack coordination, thus yielding a large amount of taxonomic and distributional data that are scattered through time in different bibliographies and natural history collections (Sastre & Lobo, 2009). In particular, the oldest records have been generally inaccessible to many researchers, leading to taxonomic errors and redundant work. Although this problem appears difficult to solve, we can optimise our future efforts by analysing how the taxonomic process occurred in the past. First, we can assess the factors that might govern the species description process (i.e. alpha taxonomy, Mayr, 1969). Species that have been described earlier share biogeographical and morphological attributes compared to species that were described later. Therefore, we can anticipate the characteristics of species yet-to-be-described to guide our search for them (Gaston & Blackburn, 1994; Diniz-Filho *et al.*, 2005). Second, we can assess the factors that influence whether species are prone to be independently (and mistakenly) described under different names. The refinement of taxonomies, such as by removing synonyms, usually occurs during the second stage of the taxonomic process (i.e. beta taxonomy, Mayr, 1969) when entire taxonomic groups are subject to revisions. This stage can be optimised by assessing the characteristics of a species that influence the probability of generating synonyms, potentially allowing to allocate more beta taxonomic effort in particular subsets. To our knowledge, assessment of species characteristics that affect the probability of redundant description (i.e. generating synonyms) has only been performed for a wasp family (Baselga *et al.*, 2010), so it is very difficult to make generalisations about the process. Therefore, new studies on this topic are needed in order to optimise our beta-taxonomic effort.

The large number of potential factors and complexity of their interactions makes it difficult to predict the probability that a given species is discovered (Jiménez-Valverde & Ortuño, 2007). Nevertheless, it seems that species described earlier may share the following characteristics: conspicuousness (Gaston, 1991; Gaston & Blackburn, 1994; Collen

et al., 2004; Diniz-Filho *et al.*, 2005); have a diurnal nature (Collen *et al.*, 2004); live in accessible sites, for example, near human populations (Diniz-Filho *et al.*, 2005), taxonomists' homes (Dennis & Thomas, 2000) or research centres (Sánchez-Fernández *et al.*, 2008); and have wide distributions (Cabrero-Sañudo & Lobo, 2003; Collen *et al.*, 2004; Gibbons *et al.*, 2005; Guil & Cabrero-Sañudo, 2007). Regarding the probability of being redundantly described, species described earlier and with widespread ranges seem to have more synonyms, as shown among eulimid wasps (Baselga *et al.*, 2010), but the scarcity of studies on this topic precludes making any general statements. Even regarding the cited paper (Baselga *et al.*, 2010), it must be clarified whether the relationship between the number of synonyms and the year of description is the result of an increment of taxonomic accuracy with time or the mere effect of available time to accumulate synonyms under a constant rate of taxonomic errors. Therefore, if we want to improve the efficiency of the taxonomic process to inventory biodiversity, we must identify the regions where future surveys should be conducted and the characteristics of undetected synonyms that hamper taxonomic knowledge.

Here, we aimed to study the historical patterns in taxonomy of Chrysomelidae (Coleoptera) in Continental Europe. The Palearctic beetles in this region have been comparatively more studied by professional and amateur entomologists than those in other parts of the World. Current knowledge on Palearctic beetles, however, is still lagging behind other groups, such as vertebrates and plants (Löbl & Smetana, 2010). The family Chrysomelidae is a hyperdiverse group with more than 2000 genera (Seeno & Wilcox, 1982) and more than 37 000 described species (Jolivet & Hawkeswood, 1995) distributed across all continents. The number of the European Chrysomelidae already described is above 2000 species according to Fauna Europaea (Audisio, 2013).

We analysed which species' characteristics are related to their probability of discovery and probability of redundant description. To estimate these probabilities, we use the year of first description and the number of synonyms of the species as proxies, assuming that the higher the probability of description, the earlier the species has been described and, independently, the higher the probability of being redundantly described as a synonym, the larger the number of synonyms. These proxies were regressed against a set of species characteristics that could potentially account for their variation: (i) spatial variables describing the species range, (ii) distribution extent, (iii) beetle size, and (iv) year of description (only for the models with number of synonyms as response variable).

Methods

Data

The list of European Chrysomelidae species and subspecies was obtained from Fauna Europaea ($n = 2249$).

This list details the name, year of description and distribution (presence/absence) of the species in European countries or territories (e.g. European Russia is divided into three separate territories: northern, central and southern; the Republics of Serbia and Montenegro are included as a single unit; and only the European portion of Turkey is considered). For this study, only continental Europe is considered and Bruchinae were discarded because they have been traditionally included in a different family based on their specialised ecology (seed predators) and morphology. Bruchinae have been traditionally studied by different specialists, and their descriptions and taxonomic revisions have been subject to a historical process that is unique to this subfamily. Thus, the total number of species used for this study is 1599 (Table 1).

The number of synonyms of each species was retrieved from monographic revisions and catalogues of the group (Doguet, 1994; Gruev & Döberl, 1997; Löbl & Smetana, 2010). Average body length of the species was computed from the range of values (minimum and maximum size) available in the literature (e.g. Doguet, 1994; Petitpierre, 2000; Warchalowski, 2003; Konstantinov *et al.*, 2011) as the arithmetic mean of the range of values. For species with no information on their body size, the mean body size of the species within the same genus was used to avoid reducing the data set. This was only conducted in genera with data available for more than 50% of the species.

Species maximum and minimum latitudes and longitudes were estimated from the boundaries of the European territories where they are present. The latitudinal and longitudinal ranges were computed as the maximum minus the minimum latitudes and longitudes respectively. Mean latitude and longitude were computed as the arithmetic mean of the corresponding maximum and minimum values. The area of the species distribution, that is, the sum of the areas of the countries where the species are present, was computed using 1:110 m Cultural Vectors from www.naturalearthdata.com in R (R Core Team, 2015).

Statistical analyses

To assess the probability of a species being: (i) discovered or (ii) redundantly described, we used two proxies:

Table 1. Total number of studied species, valid names with associated synonyms and total number of synonyms for the full study period (1758–2008) and for the first half (1758–1883) of the study period.

	Full study period	1758–1883 (%)
Total species	1599	998 (62.4)
Valid species names with associate synonyms	826 (51.7%)	709 (71.0)
Total synonyms	2752	1784 (64.8)

the year of description and the number of synonyms respectively. Therefore, we built regression models wherein the (i) year of description and (ii) number of synonyms [normalised as $\log_{10}(n + 1)$] were used as response variables. Several variable sets were used as potential predictors: spatial variables (maximum, minimum and mean latitude and longitude), distribution extent (area and latitudinal and longitudinal ranges) and beetle size (body length). Additionally, in the regression of the number of synonyms, the year of description was used as a predictor, and we also assessed whether the relationship between the number of synonyms and the year of description changes over time. To do this, we extracted subsets of the data set by removing the species described in the last 50, 75 and 100 years and conducted an independent regression on each subset. We sought to understand whether the observed relationship might be driven by an improved taxonomic efficiency over time (i.e. removing recent species would not change the relationship) or, alternatively, by the fact that recent species have not been subject to monographic revisions (beta taxonomy). In this case, the number of synonyms in species subject to revisions would be roughly constant and the non-revised species, with no synonyms, would drive the negative relationship (which would disappear after removing the species described in the most recent years).

Some assumptions of linear models may be violated if data points are not independent. If the residuals of the linear models show phylogenetic signals, then they are correlated. To account for this, we performed phylogenetic generalised least-squares regressions (PGLS; Freckleton *et al.*, 2002) using the function *gls* in R package NLME (Pinheiro *et al.*, 2016) and providing a correlation structure obtained using the function *corPagel* in package APE (Paradis *et al.*, 2004) and estimating the lambda parameter by maximum likelihood. We constructed a phylogeny of Chrysomelidae using a published phylogeny of Iberian leaf beetles (Gómez-Rodríguez *et al.*, 2015, appendix S2b in their supplementary material) as a base, and including species and subspecies for which genetic data were not available by using R package PASTIS (Thomas *et al.*, 2013) and MrBayes (Ronquist *et al.*, 2012). We only added species and subspecies belonging to genera already present in the tree, resulting in 1409 species and subspecies being included in subsequent analyses.

To identify significant variables, univariate regression models were computed for each predictor variable. Within each variable set (spatial position, extent, size and, in the case of the model for number of synonyms, the year of description), significant predictors in univariate regressions were subject to a variable selection forward procedure based on Akaike's information criterion in order to obtain the most parsimonious regression model. In the case of the spatial variables, we excluded latitudinal and longitudinal means from the set because they are linearly related to the maxima and minima and that would cause problems when fitting the PGLS models. Final models for each variable set were used in a variance partitioning

analysis in order to identify the unique and joint contributions of each set of variables.

Results

Overall patterns

The Fauna Europaea list comprises 2249 species, of which 1599 were used in our analyses. In this subset, the year of description ranges from 1758, when 55 species were described, to 2008, when *Cryptocephalus (Cryptocephalus) halleri* Costesséque was described (Fig. 1a). The rate of description was highest between 1840 and 1900, with maximum values observed in the 1850s. Notably, a reduction in the description rate is observed during the period of the Second World War and in recent years (Fig. 1a). Body length ranged between 1.05 mm (*Aphthona venustula attica* Weise, 1890) and 17.5 mm (*Timarcha pimeloides* Herrich-Schäffer, 1838). Approximately 52% of valid species names had associated synonyms (Table 1), which are more numerous in species described during the first half of the study period (Fig. 1b). In 1883 (mid-way of our study period), the total described species was 998, which was more than half of the total, and 709 of them have associated synonyms (Table 1; Fig. 1b). Therefore, more than 85% of species names that have associated synonyms were described before 1883.

Year of description

All predictors showed a statistically significant relationship with the year of description of a species, except for

body length and mean longitude (Table 2). Area of distribution had a negative relationship with the year of description of a species, and it explained the largest amount of variance in the PGLS models (pseudo $R^2 = 0.34$, $P < 0.001$). Among the statistically significant predictors, the lowest value of explained variance was observed for minimum latitude ($R^2 = 0.04$, $P < 0.001$; Table 2).

The spatial model, after a forward selection of variables, included maximum and minimum longitude, and maximum latitude ($R^2 = 0.36$, $P < 0.001$). The distribution extent model (area) explained a similar amount of variation ($R^2 = 0.34$, $P < 0.001$). The final model, used for variance partitioning, explained 36% (Table 2) of the variance in the year of description. The unique effect of each variable set was thus almost negligible (<3%), as most of the variability was jointly explained by both the spatial and extent variable sets (34%, see Fig. 2).

Number of synonyms

For the year of description, all predictors showed a statistically significant relationship with the number of synonyms, except for body length (Table 3). Year of description had a negative relationship and explained the largest amount of variance in PGLS models (pseudo $R^2 = 0.43$, $P < 0.001$). Among significant predictors, the minimum value of explained variance was observed for mean longitude (pseudo $R^2 = 0.003$, $P = 0.0211$; Table 3).

The spatial model included maximum and minimum latitude (pseudo $R^2 = 0.28$, $P < 0.001$). The distribution extent model (area) had a similar explanatory power (pseudo $R^2 = 0.29$, $P < 0.001$). The final model explained

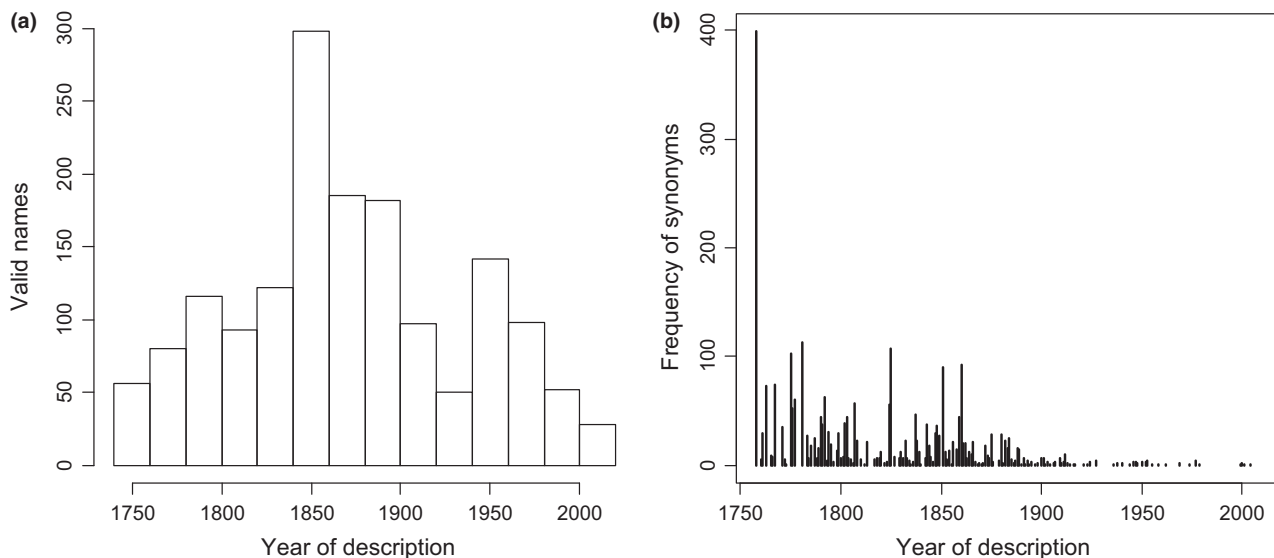


Fig. 1. Histograms representing the number of valid species described during the study period (a) and the total number of synonyms associated with all species described in each year during the study period (b).

Table 2. Phylogenetic generalised least-squares regressions between the year of description of valid names and the predictor variables used in this study.

	Pseudo R^2	P-value
Univariate models		
Spatial variables		
Maximum latitude	0.2727	<0.001
Minimum latitude	0.0389	<0.001
Mean latitude	0.1841	<0.001
Maximum longitude	0.1338	<0.001
Minimum longitude	0.0928	<0.001
Mean longitude	0.0003	0.2354
Distribution extent		
Area of distribution	0.3383	<0.001
Latitudinal range	0.3286	<0.001
Longitudinal range	0.3461	<0.001
Beetle size		
Body length	0.0006	0.1723
Multivariate models		
Spatial variables		
Maximum latitude + maximum longitude + minimum longitude	0.3643	<0.001
Distribution extent		
Area	0.3381	<0.001
Final model		
Maximum latitude + maximum longitude + minimum longitude + area	0.3638	<0.001

46% of the variance in the number of synonyms (Table 3). The year of description showed the largest unique effect (17%), whereas the unique effect of distribution extent and spatial variables was almost negligible (<1% in both cases). Most of the variability was jointly explained by the three variable sets (24%, see Fig. 2).

The relationship between the number of synonyms and the year of description was similar after removing the last years of the study, that is, explained variance decreases

from 43% (full data set) to 40%, 37% and 33% ($P < 0.001$ in all cases) when the last 50, 75 and 100 years were removed respectively (Table 4).

Discussion

Our results show that the probability of description of European Chrysomelidae is determined mostly by the characteristics of the species distributions, but not by their body size, unlike other groups (Gaston, 1991; Gaston & Blackburn, 1994; Medellín & Soberón, 1999). In short, we found that species with wide distributions and living in northern Europe have been described before. Additionally, in leaf beetle species the probability of being redundantly described is mainly related to the year of description, and to the location and distribution extent, but it is not related to the body size. Hence, species with more synonyms were described earlier and have broader distributions. An extrapolation of these trends into the future suggests that the species of Chrysomelidae that are still undescribed have narrow distributions and live in southern Europe. Regarding synonyms, our results suggest that the error rate in alpha-taxonomic studies for generating redundant descriptions has regularly decreased over time.

Probability of description

The relevance of species distributions' attributes in determining the probability of describing good species in leaf beetles is similar to that of other taxa. In turn, body size was a poor predictor in European Chrysomelidae, despite being a relevant predictor of the probability of description in other groups, such as British beetles (Gaston, 1991), birds worldwide (Gaston & Blackburn, 1994), North American butterflies (Gaston *et al.*, 1995a,b), anurans in Brazilian Cerrado (Diniz-Filho *et al.*, 2005) and

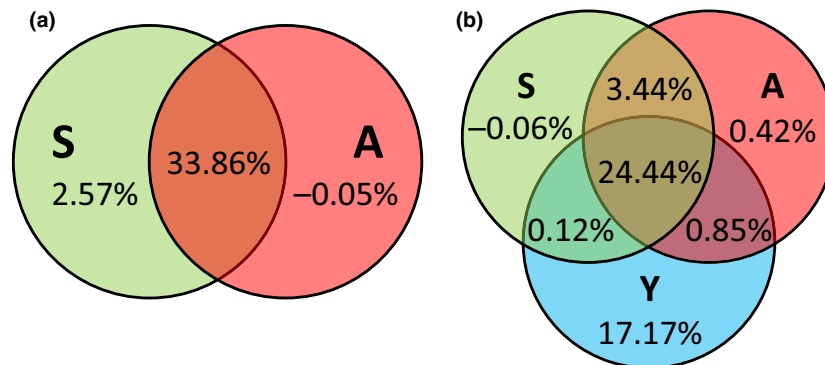


Fig. 2. Unique and shared variance explained in multiple regression models by spatial variables (S), extent of distribution area (A) and year of description (Y). Results are shown for the year of description (a) and number of synonyms as response variables (b). [Colour figure can be viewed at wileyonlinelibrary.com]

Table 3. Phylogenetic generalised least-squares regressions between the number of synonyms and the predictor variables used in this study.

		R^2	P -value
Univariate models			
Spatial variables			
Maximum latitude		0.2260	<0.001
Minimum latitude		0.0345	<0.001
Mean latitude		0.1508	<0.001
Maximum longitude		0.1340	<0.001
Minimum longitude		0.0650	<0.001
Mean longitude		0.0031	0.0211
Distribution extent			
Area of distribution		0.2914	<0.001
Latitudinal range		0.2741	<0.001
Longitudinal range		0.2934	<0.001
Beetle size			
Body length		-0.0005	0.6205
Year of description			
Year of description		0.4256	<0.001
Multivariate models			
Spatial variables	Maximum latitude + minimum latitude	0.2793	<0.001
Distribution extent	Area	0.2914	<0.001
Year of description	Year of description	0.4258	<0.001
Spatial variables + year of description	Maximum latitude + minimum latitude + year of description	0.4596	<0.001
Spatial variables + distribution extent	Maximum latitude + minimum latitude + area	0.2920	<0.001
Year of description + distribution extent	Year of description + area	0.4643	<0.001
Final model	Maximum latitude + minimum latitude + area + year of description	0.4637	<0.001

birds, butterflies and orchids in the Western Ghats (Aravind *et al.*, 2007). In this study, the observed lack of explanatory power of body size may be related to the sampling method used for leaf beetles, which is normally unspecific (often sweeping host plants to collect all the leaf beetles present). As a result, leaf beetles are caught independent of conspicuousness. This same lack of relevance for body size in determining the probability of description has also been observed in other taxa, such as American oscine passerine birds (Blackburn & Gaston, 1995), Australian scarab beetles (Allsopp, 1997), herpetofauna of North America (Reed & Boback, 2002), dung beetle species (Cabrero-Sañudo & Lobo, 2003) and Western Palearctic *Aphthona* species (Baselga *et al.*, 2007). The cited studies do not discuss the reasons for this lack

Table 4. Phylogenetic generalised least-squares regressions between the number of synonyms and the year of description.

	Pseudo R^2	P -value
Years removed		
No year removed (full data set)	0.4256	<0.001
Last 50 years	0.3953	<0.001
Last 75 years	0.3670	<0.001
Last 100 years	0.3346	<0.001

Results are provided for the full data set (full range of the year of description) and for subsets of this data set where some periods have been removed (the last 50 years, the last 75 years and the last 100 years).

of relationship between body size and probability of description, but we can speculate that, as discussed previously for European leaf beetles, body size is not important in determining the probability of description when the same sampling method is used for all species in the analysed group. For example, all dung beetles are sampled using traps baited with excrements, so even if large body size differences exist, it does not impact the probability of detection. In contrast, when all beetle families have been assessed together (as in Gaston, 1991), because different sampling methods are applied for different families and body size differences among families exist, body size does affect the probability of description.

Among the geographical characteristics of the distribution, the size of the distribution area is a major determinant of the probability of description. In other words, widely distributed species were described earlier and have more synonyms than species with restricted distributions. The size of the distribution area is a significant predictor of the likelihood of description for many groups (Blackburn & Gaston, 1995; Allsopp, 1997; Cabrero-Sañudo & Lobo, 2003; Collen *et al.*, 2004), showing that widespread and common species are more likely to be found and, thus, discovered first. Our measure of range size is in fact a surrogate (i.e. the sum of areas of countries where the species are present) and not an accurate quantification of the area occupied by the species. The latter information is not available for the vast majority of the species, but our surrogate measure should capture the major trends in the

data. An alternative surrogate (i.e. the number of countries where a species is present) is tightly correlated with the one we used ($r^2 = 0.92$), so our results should be robust to using this alternative measure of range size. The effect of area on the probability of description also shows a geographically structured pattern. Species with restricted distributions in Europe are usually found in the southern latitudes (Lumaret & Lobo, 1996; Baselga, 2008). The southern latitudes have often been refuges during glaciations, while northern zoological assemblages are mainly composed of postglacial colonisers (Taberlet *et al.*, 1998; Schmitt & Krauss, 2004; Sommer & Benecke, 2005). Thus, species with northern distributions usually have broader distributions than southern species (Boucher-Lalonde *et al.*, 2016). Nevertheless, given that species are described by people, the geographical pattern in the probability of description might be related not only to the species' biogeography, but also to anthropogenic factors. Taxonomists have traditionally lived in countries of Northern Europe (Cabreró-Sañudo & Lobo, 2003), so it is expected that northern and broad-ranging species were described first, as we have shown.

Probability of accumulating synonyms

Regarding the probability of redundant descriptions, our results show that body size was not a good predictor. The probability of being redundantly described increases with the increment of encounters between the species and the specialists. Thus, as for the probability of description, species with larger distribution ranges that include the northern latitudes tend to accumulate more synonyms. This pattern is also related to time. We stress that the trends in the accumulation of synonyms and hence, in the efficiency of taxonomy, have been rarely studied. Gaston *et al.* (1995a) reported a correlation between the number of synonyms and the year of description, but the only previous detailed study showing that several factors affect the probability of redundant descriptions was, to our knowledge, devoted to eupelmid wasps (Baselga *et al.*, 2010). Therefore, it is difficult to establish generalities yet, although our results also show that the probability of being redundantly described is related to the year of description, as we found that species with more synonyms were described before the 19th century (85% of species with associated synonyms accumulated during the first half of the studied time range; Table 1; Fig. 2). This can be interpreted in two ways. On the one hand, it could be argued that communication between researchers in the past was less efficient (compared with today's flux of information), increasing the probability that more than one name was given to a species (i.e. the effectiveness of alpha taxonomy could be worse in the past, Baselga *et al.*, 2010). Therefore, present species descriptions should be more accurate currently, and we should not expect current descriptions to be synonymised in the future. An alternative interpretation would imply that an imperfect

taxonomic accuracy might remain constant over time, so that species described earlier have more synonyms just because more time elapsed, allowing the detection of synonyms in subsequent taxonomic revisions (beta taxonomy). This would imply that taxonomic descriptions are not more accurate today. Instead, we would not be detecting redundant descriptions of recently described species simply because recent species have not been included in taxonomic revisions yet (i.e. lack of beta taxonomy for recently described species, Baselga *et al.*, 2010). Here, we assessed the likelihood of these two scenarios (higher error rate in the past vs. undetected synonyms in recent years) for the first time by analysing whether the relationship between the number of synonyms and the year of description remains the same over time (i.e. we removed the last 50, 75 and 100 years of the study for the statistical analyses). If the relationship was driven by the undetected synonyms of recent species, we would expect a strong decrease in the R^2 of this relationship when removing the last years (because the 100-year-old species would have had enough time to accumulate synonyms through subsequent taxonomic revisions). Although R^2 does decrease, the differences are small, suggesting that the larger number of synonyms in the past is largely due to a higher error rate in the first years of the taxonomic process. In other words, the scarcity of synonyms in recent years is apparently due to an increase in the accuracy of alpha taxonomy, although we cannot discard the existence of a fraction of undetected synonyms that may be detected by future beta-taxonomic efforts.

Current knowledge on the biodiversity of European Chrysomelidae is incomplete, as new species are routinely described. It is also likely that some synonyms remain to be detected in the future. This implies that biodiversity patterns and our inferences might not be perfectly accurate (Hortal *et al.*, 2015). In this sense, identifying biases and gaps in the taxonomic knowledge could facilitate and stimulate taxonomic research (Brûlé & Touroult, 2014). In European leaf beetles, broadly distributed and northern species have been described before, so they accumulate more synonyms. The extrapolation of past taxonomic trends to the future allows us to make recommendations about where to allocate alpha- and beta-taxonomy efforts. Alpha-taxonomy (i.e. the search for new species) should be reinforced in Southern Europe, particularly in isolated habitats where new small-ranging species could still be discovered. In turn, regarding beta-taxonomy, our results suggest that although it seems likely that most redundant descriptions of leaf beetle species have already been detected, future beta-taxonomic efforts (i.e. taxonomic revisions) could still detect new synonyms (i.e. junior names that correspond to older, widespread species).

Acknowledgements

The authors received the funding from the Spanish Ministry of Economy and Competitiveness and the European

Regional Development Fund (ERDF) through grants CGL2013-43350-P and CGL2016-76637-P.

References

- Allsopp, P.G. (1997) Probability of describing an Australian scarab beetle: influence of body size and distribution. *Journal of Biogeography*, **24**, 717–724.
- Aravind, N.A., Tambat, B., Ravikanth, G., Ganeshiah, K.N. & Uma Shaanker, R. (2007) Patterns of species discovery in the Western Ghats, a megadiversity hot spot in India. *Journal of Biosciences*, **3**, 781–790.
- Audisio, P. (2013) Fauna Europaea: Chrysomelidae (2013) Fauna Europaea version 2.6.2. <http://www.fauna-eu.org> 29th July 2015.
- Baselga, A., Hortal, J., Jiménez-Valverde, A., Gómez, J.F. & Lobo, J. (2007) Which leaf beetles have not yet been described? Determinants of the description of Western Palearctic *Aphthona* species (Coleoptera: Chrysomelidae). *Biodiversity and Conservation*, **16**, 1409–1421.
- Baselga, A. (2008) Determinants of species richness, endemism and turnover in European longhorn beetles. *Ecography*, **31**, 263–271.
- Baselga, A., Lobo, J.M., Hortal, J., Jiménez-Valverde, A. & Gómez, J.F. (2010) Assessing *alpha* and *beta* taxonomy in eupelmid wasps: determinants of the probability of describing good species and synonyms. *Journal of Zoological Systematics and Evolutionary*, **48**, 40–49.
- Blackburn, T.M. & Gaston, K.J. (1995) What determines the probability of discovering a species? A study of South American oscine passerine birds. *Journal of Biogeography*, **22**, 7–14.
- Boucher-Lalonde, V., Morin, A. & Currie, D.J. (2016) Can the richness-climate relationship be explained by systematic variations in how individual species' ranges relate to climate? *Global Ecology and Biogeography*, **25**, 527–539.
- Brito, D. (2010) Overcoming the Linnean shortfall: data deficiency and biological survey priorities. *Basic and Applied Ecology*, **11**, 709–7013.
- Brown, J.H. & Lomolino, M.V. (1998) *Biogeography*, 2nd edn. Sinauer Associates Inc, Sunderland, Massachusetts.
- Brülé, S. & Tourout, J. (2014) Insects of French Guiana: a baseline for diversity and taxonomic effort. *ZooKeys*, **434**, 111–130.
- Cabrero-Sañudo, F.J. & Lobo, J.M. (2003) Estimating the number of species not yet described and their characteristics: the case of Western Palearctic dung beetle species (Coleoptera, Scarabaeoidea). *Biodiversity and Conservation*, **12**, 147–166.
- Collen, B., Purvis, A. & Gittleman, J.L. (2004) Biological correlates of description date in carnivores and primates. *Global Ecology and Biogeography*, **13**, 459–467.
- Cotterill, F.P.D. (1995) Systematics, biological knowledge and environmental conservation. *Biodiversity and Conservation*, **4**, 183–205.
- Cracraft, J. (1996) Systematics, biodiversity science, and the conservation of the Earth's biota. *Verhandlungen der Deutschen Zoologischen Gesellschaft*, **89**, 41–47.
- Dennis, R.L.H. & Thomas, C.D. (2000) Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. *Journal of Insect Conservation*, **4**, 73–77.
- Diniz-Filho, J.A.F., De Marco, P.J.R. & Hawkins, B.A. (2010) Defying the curse of ignorance: perspectives in insect macroecology and conservation biogeography. *Insect Conservation and Diversity*, **3**, 172–179.
- Diniz-Filho, J.A.F., Pereira Bastos, R., Rangel, T.F.L.V.B., Bini, L.M., Carvalho, P. & Silva, R. (2005) Macroecological correlates and spatial patterns of anuran description dates in the Brazilian Cerrado. *Global Ecology and Biogeography*, **14**, 469–477.
- Doguet, S. (1994) *Coléoptères Chrysomelidae Vol. 2 Alticinae*. Faune de France 80. Federation Francaise des Societes de Sciences Naturelles, Paris, France.
- Freckleton, F.P., Harvey, P.H. & Pagel, M. (2002) Phylogenetic analysis and comparative data: a test and review of evidence. *The American Naturalist*, **160**, 712–726.
- Gaston, K.J. (1991) Body size and probability of description: the beetle fauna of Britain. *Ecological Entomology*, **16**, 505–508.
- Gaston, K.J. & Blackburn, T.M. (1994) Are newly described bird species small-bodied? *Biodiversity Letters*, **2**, 16–20.
- Gaston, K.J., Blackburn, T.M. & Loder, N. (1995a) Which species are described first? The case of North American butterflies. *Biodiversity and Conservation*, **4**, 119–127.
- Gaston, K.J. & May, R.M. (1992) Taxonomy of taxonomists. *Nature*, **356**, 281–282.
- Gaston, K.J. & Mound, L. (1993) Taxonomy, hypothesis testing and the biodiversity crisis. *Proceedings of the Royal Society of London: Biological Sciences*, **251**, 139–142.
- Gaston, K.J., Scoble, M.J. & Crook, A. (1995b) Patterns in species description: a case study using the Geometridae (Lepidoptera). *Biological Journal of the Linnean Society*, **55**, 225–237.
- Gibbons, M.J., Richardson, A.J., Angel, M.V., Buecher, E., Esnal, G., Fernandez, M.A., 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.
- Gómez-Rodríguez, C., Crampton-Platt, A., Timmermans, M.J.T.N., Baselga, A. & Vogler, A.P. (2015) Validating the power of mitochondrial metagenomics for community ecology and phylogenetics of complex assemblages. *Methods in Ecology and Evolution*, **6**, 883–894.
- Gruev, B. & Döberl, M. (1997) General distribution of the flea beetles in the Palearctic Subregion (Coleoptera, Chrysomelidae: Alticinae). *Scopolia: Journal of the Slovenian Museum of Natural History*, **37**, 1–496.
- Guil, N. & Cabrero-Sañudo, F.J. (2007) Analysis of the species description process for a little known invertebrate group: the limnoterrestrial tardigrades (Bilateria, Tardigrada). *Biodiversity and Conservation*, **16**, 1063–1086.
- Hammond, P. (1992) Species inventory. *Global Biodiversity* (ed. by B. Groombridge), pp. 17–39. World Conservation Monitoring Centre, Chapman and Hall, London, UK.
- Hebert, P.D.N., Cywinska, A., Ball, S.L. & DeWaard, J.R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, 313–321.
- Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015) Seven shortfalls that beset large-scale knowledge of biodiversity. *The Annual Review of Ecology, Evolution and Systematics*, **46**, 523–549.
- Jiménez-Valverde, A. & Ortuño, V.M. (2007) The history of endemic Iberian ground beetle description (Insecta, Coleoptera, Carabidae): which species were described first? *Acta Oecologica*, **31**, 13–31.
- Jolivet, P. & Hawkeswood, T.J. (1995) *Host-Plants of Chrysomelidae of the World*. Backhuys Publishers, Leiden, The Netherlands.

- Konstantinov, A.S., Baselga, A., Grebennikov, V.V., Prena, J. & Lingafelter, S.W. (2011) *Revision of the Palearctic Chaetocnema species (Coleoptera: Chrysomelidae: Galerucinae: Alticini)*. Pensoft Publishers, Sofia, Bulgaria.
- Löbl, I. & Smetana, A. (2010) *Catalogue of Palearctic Coleoptera, Chrysomeloidea*, Vol. 6. Apollo Books, Stenstrup, Denmark.
- Lumaret, J.P. & Lobo, J.M. (1996) Geographic distribution of endemic dung beetles (Coleoptera, Scarabaeoidea) in the Western Palearctic Region. *Biodiversity Letters*, **3**, 192–199.
- May, R.M. (1988) How many species are there on Earth? *Science, New Series*, **241**, 1441–1449.
- May, R.M. (2010) Tropical arthropod species, more or less? *Science*, **329**, 41–42.
- Mayr, E. (1969) *Principles of Systematic Zoology*. McGraw-Hill, New York City, New York.
- Medellín, R.A. & Soberón, J. (1999) Predictions of mammal diversity of four land masses. *Conservation Biology*, **13**, 143–149.
- Morell, V. (1996) New mammals discovered by biology's new explorers. *Science*, **273**, 1491.
- Odegaard, F. (2000) How many species of arthropods? Erwin's estimate revised. *Biological Journal of the Linnean Society*, **71**, 583–597.
- Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.
- Patterson, B.D. (1994) Accumulating knowledge on the dimensions of biodiversity: systematic perspectives on Neotropical mammals. *Biodiversity Letters*, **2**, 79–86.
- Petitpierre, E. (2000) Coleoptera, Chrysomelidae I. *Fauna Ibérica* Vol. 13 Museo Nacional de Ciencias Naturales CSIC, Madrid, Spain.
- Pimm, S.L., Russell, G.J., Gittleman, J.L. & Brooks, T.M. (1995) The future of biodiversity. *Science*, **269**, 347–350.
- Pinheiro, J., Bates, D., DebRoy, S. & Sarkar, D. & R Core Team (2016) nlme: linear and nonlinear mixed effects models. R package version 3.1-128. <<http://CRAN.R-project.org/package=nlme>> 14th November 2016.
- Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D. & Vogler, A.P. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595–609.
- R Core Team (2015) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <<https://www.R-project.org/>> 20th October 2016.
- Reed, R.N. & Boback, S.M. (2002) Does body size predict dates of species description among North American and Australian reptiles and amphibians? *Global Ecology and Biogeography*, **11**, 41–47.
- Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, L.D., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A. & Huelsenbeck, J.P. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.
- Sánchez-Fernández, D., Lobo, J.M., Abellán, P., Ribera, I. & Millán, A. (2008) Bias in freshwater biodiversity sampling: the case of Iberian water beetles. *Diversity and Distributions*, **14**, 754–762.
- Sastre, P. & Lobo, J.M. (2009) Taxonomist survey biases and the unveiling of biodiversity patterns. *Biological Conservation*, **142**, 462–467.
- Schmitt, T. & Krauss, J. (2004) Reconstruction of the colonization route from glacial refugium to the northern distribution range of the European butterfly *Polyommatus coridon* (Lepidoptera: Lycaenidae). *Diversity and Distributions*, **10**, 271–274.
- Seeno, N.T. & Wilcox, J.A. (1982) Leaf beetle genera (Coleoptera: Chrysomelidae). *Entomography*, **1**, 1–221.
- Sommer, R.S. & Benecke, N. (2005) The recolonization of Europe by brown bears *Ursus arctos* Linnaeus, 1758 after the last glacial maximum. *Mammal Review*, **35**, 156–164.
- Stork, N.E. (1997) Measuring global biodiversity and its decline. *Biodiversity II: Understanding and Protecting Our Biological Resources* (ed. by M.L. Reaka-Kudla, D.E. Wilson and E.O. Wilson), pp. 41–68. Joseph Henry Press, Washington, District of Columbia.
- Taberlet, P., Fumagalli, L., Wust-Saucy, A.G. & Cosson, J.F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.
- Thomas, G.H., Hartmann, K., Jetz, W., Joy, J.B., Mimoto, A. & Mooers, A.O. (2013) PASTIS: an R package to facilitate phylogenetic assembly with soft taxonomy inferences. *Methods in Ecology and Evolution*, **4**, 1011–1017.
- Warchalowski, A. (2003) *Chrysomelidae, the Leaf-Beetles of Europe and the Mediterranean Area*. Natura optima dux Foundation, Warsaw, Poland.
- Whittaker, R.J., Araújo, M.B., Jepson, P., Ladle, R.J., Watson, J.E.M. & Willis, K.J. (2005) Conservation Biogeography: assessment and prospect. *Diversity and Distributions*, **11**, 3–23.
- Williams, P.H. (1998) An annotated checklist of bumble bees with an analysis of patterns of description (Hymenoptera: Apidae, Bombini). *Bulletin of the Natural History Museum (Entomology)*, **67**, 79–152.

Accepted 12 April 2017

First published online 15 May 2017

Editor: Simon Leather

Associate editor: Gavin Broad