



Estimating the number of species not yet described and their characteristics: the case of Western Palearctic dung beetle species (Coleoptera, Scarabaeoidea)

F.J. CABRERO-SAÑUDO* and J.M. LOBO

*Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC), C/ José Gutiérrez Abascal 2, E-28006 Madrid, Spain; *Author for correspondence (e-mail: fcabrero@mncn.csic.es; fax: +34-91-5645078)*

Received 20 June 2001; accepted in revised form 8 January 2002

Key words: Geographic location, Geographic range size, Probability of discovery, Species description rate, Variation partitioning

Abstract. To determine the degree of completion of Western Palearctic dung beetle species inventory, as well as to ascertain some of the main geographical and morphological characteristics associated with the probability of new species description, the available taxonomic and biogeographic information of the three Scarabaeoidea families that are associated with dung was analyzed. An asymptotic fit of the cumulative number of described species is used to estimate the approximate number of species not yet described. Variation partitioning and hierarchical decomposition techniques were used to explore the influence of body size, geographical range size and geographical location variables on the process of species description. Results indicate that the inventories of Western Palearctic Scarabaeidae and Geotrupidae species are almost complete, but that around 16% of Aphodiidae species have yet to be described. As the joint effect of geographical distribution and location is the main determinant of the species description rate in this family, it was predicted that new dung beetle species would be found in smaller areas of the southern and eastern Western Palearctic region. The species body size accounted for a negligible fraction in the variation of the year of description. Both the geographical range size and the geographical location are essential variables negatively correlated with the year of species description, whose pure and combined effect can account for high percentages in the year of description variation (from 41 to 56%). The location of most of the taxonomic workforce in areas where broader-ranged generalist species prevail in the assemblages is the probable reason for this pattern.

Abbreviations: *a* – pure effect of body size variation; AIC – Akaike information criterion, *b* – pure effect of geographic range size, BS – body size, *c* – pure effect of geographic location, *d* – combined variation due to the joint effect of body size and geographic range size components, *e* – combined variation due to the joint effect of body size and geographical location components, *f* – combined variation due to the joint effect of geographical range size and geographic location components, *g* – combined variation due to the joint effect of the three variables, GL – geographic location; GLM – Generalized Linear Models, GRS – geographic range size, U – variation not explained by independent variables included in the analysis.

Introduction

During the last few centuries and particularly more recently, the Earth's biota has

been suffering the hitherto unsuspected repercussions of the uses and abuses of man, culminating in the extinction of various species (Illies 1983; Gaston 1991a; Gaston and Mound 1993; Mayr 1993). The reality of these repercussions has been demonstrated (Heywood 1995; Lawton and May 1995; Hughes et al. 1997; Heard and Mooers 2000; Purvis et al. 2000), although we are still ignorant about their extent (Gaston 1991a). Time and again, when faced with particular biodiversity crises, human beings have responded (Mayr 1993), although the struggle to inventory all species has proved futile (Patterson 1994) since many of them disappeared before being described.

Species description is an advantageous task, both to increase the knowledge of species (May 1988) and to aid in their conservation (Sutton and Collins 1991). However, knowledge would still be lacking, especially in the case of insects, not only on the number of species still to be discovered in a given region, but also on the factors that might govern the species description process. Another equally important problem is the decrease in resources assigned to systematics (Cotterill 1995) and the subsequent fall from the status of 'good' science (Krell 2000), precluding any guarantee of the reliability of conservation decisions (Barrowclough 1992). In order for their work to be efficiently conducted and their results cost effective, taxonomists must make it a priority to search for new species in less thoroughly prospected regions and develop new tactics in these times of crisis (Hammond 1992; Gaston 1993), since there is neither adequate time nor resources for a traditional approach to this enormous effort (Gaston 1991a).

Determining the degree of completion of species inventories in a region and estimating the geographical and morphological characteristics associated with the probability of species description can aid in establishing both efficient strategies for the discovery and location of species yet to be described. The approximate total number of species of a group and the amount of descriptive work yet to be done can be estimated by examining the growth over time of the cumulative species number as a function of the year of description (Steyskal 1965; Arnett 1967; White 1975, 1979; Frank and Curtis 1979; O'Brien and Wibmer 1979; Medellín and Soberón 1999).

On the other hand, the influence on the process of species description of body size, geographical range size and geographical location of species has been explored (Gaston 1991b, 1993, 1994, 1996; Hammond 1992; Gaston and Mound 1993; Gaston and Blackburn 1994; Patterson 1994; Blackburn and Gaston 1995; Gaston et al. 1995a; Allsopp 1997). Evidently, several other variables might influence the probability of species description (shine, coloration, population size, habitat preference, etc.) but probably have not been considered because of the greater difficulty in quantifying them. The first species attribute examined was body size (Gaston 1991b; Gaston and Blackburn 1994; Patterson 1994; Blackburn and Gaston 1995; Gaston et al. 1995a; Allsopp 1997). Assuming that there is a positive correlation between body size and the detectability of species, larger species would tend to be described before smaller species on some occasions (Gaston 1991b; Gaston et al. 1995a); newly described species would thus tend to be physically smaller than expected (Gaston and Blackburn 1994). Subsequently, species' geographical range size or

their geographical location as determinants of the probability of species description were also studied (Gaston 1993; Patterson 1994; Blackburn and Gaston 1995; Gaston et al. 1995a, b; Allsop 1997). Species with large geographic ranges tend to be described earlier; geographical range size is frequently a better predictor of the date of discovery than body size. As the rate of species description generally varies throughout the territory considered due to the concentration of faunistic exploration near taxonomists' places of residence, geographical location of species also seems to be an important variable related to the probability of species description (Patterson 1994; Allsop 1997). However, this factor has never been thoroughly studied beyond the consideration of species' occurrence in one supposed region or another, whose borders are not always clear or properly defined (e.g. latitude or longitude have not yet been used).

Estimating the influence of these variables on the process of species description is hindered by their frequent collinearity. Thus, the studies that have been carried out have been unable to quantify the true influence of these variables or the importance of their combined effects (Gaston and Blackburn 1994; Blackburn and Gaston 1995). We confront these drawbacks by applying the techniques of variation partitioning (Borcard et al. 1992; Qinghong and Bråkenhielm 1995; Anderson and Gribble 1998; Legendre and Legendre 1998) and hierarchical decomposition (MacNally 2000) in order to estimate how much of the variation of the response variable can be exclusively attributed to one independent variable, and to calculate the fraction of the variation accounted for by the joint effect of two or three of the variables considered.

The present paper analyses how the species description process has been carried out in the Western Palaearctic region for a taxonomically well-known dung beetle group (Coleoptera, Scarabaeoidea), in which three families have traditionally been included: Aphodiidae, Scarabaeidae, and Geotrupidae. Even though this is a rather well-studied species group, the total number of species included in these families is unknown. Thus, we would like to know: how many species have yet to be discovered in the Western Palaearctic region; what would be the probable characteristics of these as yet undescribed species; and where they could most likely be located within this region. Answers to these questions may facilitate a better design of the taxonomic work still to be accomplished.

Material and methods

Data collection

Taking into account the exhaustive taxonomic information on the Palaearctic species of the three Scarabaeoidea families (Baraud 1985, 1992; Dellacasa 1988a, b, c, 1991, 1995; López-Colón 1995; Hollande and Théron 1998), together with some additional data (see references in Dellacasa 1990, 1992; Král and Olexa 1996; Lumaret and Lobo 1996; Pittino 1996; Chikatunov and Pavlíček 1997; Král 1997), a total of 629 species were considered: 395 from the Aphodiidae, 162 from the

Scarabaeidae and 72 from Geotrupidae. Ignoring synonyms, only species that were currently recognized as valid taxa were considered. As the three families are considered monophyletic groups (Scholtz 1990), each one has been analysed independently in order to increase the predictive power of extrapolations (Gaston et al. 1995a). The information covers all the species described from 1758 to 1990 in the Western Palaearctic region. This region comprises North Africa from Morocco to Sinai, and Europe from the Atlantic Ocean to the Ural Mountains, including the Caspian Sea to the South and the western boundaries of Kazakhstan. Nine variables were summarized for each species: the year of description (dependent variable), body size (BS; length in mm), geographic range size (GRS), mean, maximum and minimum latitude and longitude of each species' geographical range. The last six variables allow us to quantify the geographic location (GL) of species.

The mean body size of species was provided by Lumaret and Kirk (1987), Lobo (1993) and Baraud (1992). However, for 12 species whose body size had not been specified in the literature, this was estimated as the mathematical mean of the body size of the species of the same genus or subgenus. Geographical range size scores are the six endemism classes defined by Lumaret and Lobo (1996) as the percentage that supposes the distribution area of each species in the entire Western Palaearctic region ($12 \times 10^6 \text{ km}^2$; i.e. class 1, species with a distribution area less than or equal to 0.1% of the Western Palaearctic area; class 2, between 0.1 and 0.5%; class 3, between 0.5 and 1%; class 4, between 1 and 5%; class 5, between 5 and 10%; and class 6, species distribution area greater than 10% of the total area of the region considered). Geographic location variables were estimated within an approximate 2.5° margin of error by examining a map of the localities and countries where the presence of the species was verified by the previously mentioned bibliographical references and other data from regional faunistic studies (Mackauer 1958; Carpaneto 1976, 1979, 1980, 1985; Dellacasa and Poggi 1981; Hanski and Kuusela 1983; Dellacasa 1987; Palmer et al. 1988; Lumaret 1990; Skidmore 1991; Martín-Piera and Lobo 1992; Arnone and Massa 1993; Lobo and Martín-Piera 1993; Arnone et al. 1995; Carpaneto and Piattella 1995a, b; Lumaret et al. 1996; Carpaneto et al. 1997; Ziani 1997a, b). Both maximum and minimum latitude or longitude indicates the distributional boundaries of each species. Mean latitude and mean longitude are defined as the mathematical means of these extreme values.

Statistical analysis

To estimate the number of species not yet discovered in each family, we used the logarithm of the cumulative number of species described in the successive years from 1758 to 1990. The use of cumulative curves presupposes that: (i) a finite number of species exist on Earth (Steyskal 1965; O'Brien and Wibmer 1979); (ii) the evolution of new species since Linnaeus is impossible (Frank and Curtis 1979); (iii) it is possible to approach a state in which all species are known, although new species become difficult to discover (Steyskal 1965; Arnett 1967; White 1975; Frank and Curtis 1979; Diamond 1985); and (iv) the majority of extant species have

been described, so that the species description rate falls gradually towards zero as the total number is approached (White 1979; Medellín and Soberón 1999). The logarithm of the cumulative number of described species was plotted against the year and the curve fitted to the Beta-p function selected as the best fit (Mielke and Johnson 1974), after trying to fit different types of species-accumulation functions (see Soberón and Llorente 1993; Flather 1996). The curvilinear function was fitted by the Quasi-Newton method using Statistica software (StatSoft Inc. 1999). The value of the asymptote determined by iteration is used to estimate the number of species of each family still not described. Between the different candidate asymptotic functions, the Beta-p model adequately represents this type of relationship described as:

$$y = a(1 - [1 + (x/c)^d]^{-b}).$$

To explore the multiple relationships between the year of description and the explanatory variables considered, generalized linear models (GLM; McCullagh and Nelder 1989; Crawley 1993) were used. In GLM, one dependent variable (not necessarily normally distributed) is linearly or non-linearly related to multiple independent variables via a link function that specifies the needed transformation of the predicted values. To find the best predictive model for each family, a backward stepwise procedure was employed in which both the forward entry and the backward removal of variables are used (Legendre and Legendre 1998). The resulting stepwise procedure model may not be unique, and often many similar models with equal or greater predictive power can be found (Daniel and Wood 1980; Copas 1983; Derksen and Keselman 1992; MacNally 2000). Thus, all possible 2^k regression models based on the k independent variables were computed, selecting the most appropriate models according to the Akaike information criterion (AIC; Akaike 1983; MacNally 2000). These models are compared with the stepwise regression solution in order to select the most representative final model for each family (Neter et al. 1985). Finally, the possible curvilinear relationship between the year of description and the statistically significant variables was explored in order to determine if either a linear, quadratic or cubic function of each selected variable increases the explanatory capacity of the model. The goodness of fit of the models obtained was measured by the deviance statistic, the change in deviance F -ratio tested and the reduction in deviance compared with that of a full model in which the number of parameters is equal to the total number of observations (McCullagh and Nelder 1989; Dobson 1999). The percentage of deviance explained has also been calculated for each model (see Dobson 1999). The adequacy of the derived final models was determined by detecting outliers that disproportionately influence the fit of the model, examining standard errors of coefficients, plotting residuals against the fitted values, and by plotting residuals in a normal probability plot (Nicholls 1989; Dobson 1999).

The independent variables selected for the regression analysis were further analysed to verify if they most likely influence the variation in the date of species description. Thus, the comparative importance of each explanatory variable, as well

as that of their complex inter-relationships, was measured by examining the variation partitioning among explanatory variables in the set of data. Firstly, the percentage of explained deviance for each one of the possible 2^k functions was calculated and the impact of each independent variable was estimated as the average effect of including this variable in all possible models built with the remaining variables (MacNally 2000). Subsequently, the total variation within the date of species description was decomposed among the three types of explanatory variables (i.e. BS, GRS and GL) and the percentage of explained deviance was calculated for eight different components (Figure 1): pure effect of body size variation (a), pure effect of the GRS (b), pure effect of the GL (c), combined variation due to the joint effect of body size and GRS components (d), combined variation due to the joint effect of body size and GL components (e), combined variation due to the joint effect of GRS and GL components (f), combined variation due to the joint effect of the three variables (g), and variation not explained by the independent variables included in the analysis (U). The decomposition of the variation in the date of species description among the three sets of explanatory variables was carried out by partial regression analysis (see Legendre and Legendre 1998). As explanatory variables were not mutually independent, this approach can aid in determining the amount of variation overlap and in clarifying the influence of the variables considered (Borcard et al. 1992; Qinghong and Bråkenhielm 1995; Anderson and Gribble 1998). In the process of variation decomposition, the date of species description (y) was regressed with the three types of variables together (BS, GRS and GL), which represent the total explained variation in the data set ($a + b + c + d + e + f + g$ in Figure 1). Regressing y with each one of the explanatory variables yields the variation separately attributable to BS ($a + d + e + g$), GRS ($b + d + f + g$), and GL ($c + e + f + g$). Subsequently, residuals of the regression of BS with GRS + GL variables were calculated, and y was regressed with these residuals in order to estimate the pure effect of BS variation (a). Fractions b and c were estimated in the same way after computing the regression residuals of GRS with BS + GL, and the regression residuals of GL with GRS + BS, respectively. The remaining variation fractions were computed according to two sets of equations (Borcard et al. 1992; Qinghong and Bråkenhielm 1995; Anderson and Gribble 1998), where:

$$\begin{aligned}d + e + g &= \text{BS} - a \\d + f + g &= \text{GRS} - b \\e + f + g &= \text{GL} - c\end{aligned}$$

and

$$\begin{aligned}d &= (\text{BS} + \text{GRS}) - (e + f + g) - (a + b) \\e &= (\text{BS} + \text{GL}) - (d + f + g) - (a + c) \\f &= (\text{GRS} + \text{GL}) - (d + e + g) - (b + c) \\g &= (d + e + g) - d - e = (d + f + g) - d - f = (e + f + g) - e - f.\end{aligned}$$

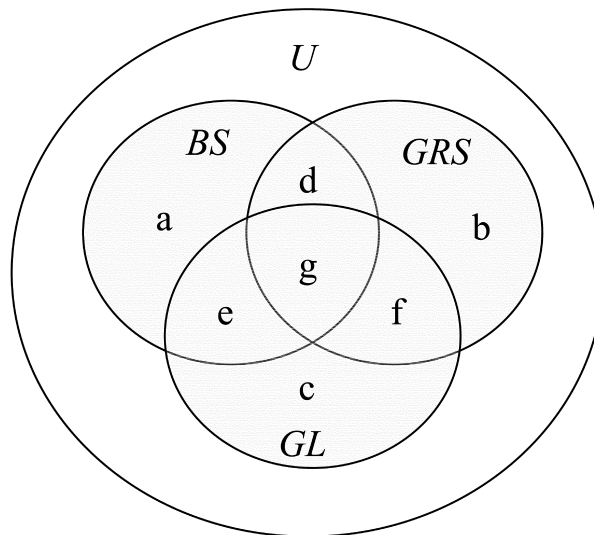


Figure 1. Diagram showing the variation in the dependent variable (the year of species description) due to three explanatory variables: BS (body size), GRS (geographic range size) and GL (geographic location). U is the unexplained variation; a , b and c are the pure effects of body size, geographic range, and location variation, respectively; while d , e and f are the combined variation due to the joint effect of body size and geographical range size, the joint effect of body size and geographic location, and the joint effect of geographical range size and geographic location, respectively; g is the combined variation due to the joint effect of the three variables.

Table 1. Descriptive statistics for each dung beetle family and results of the curvilinear fit of the logarithm of the cumulative number of described species against the year of description according to the Beta-p function (Mielke and Johnson 1974; Flather 1996).

| | Scarabaeidae | Aphodiidae | Geotrupidae |
|---|-----------------|-----------------|-----------------|
| Number of species | 162 | 395 | 72 |
| Described species per year (\pm SE) | 0.70 ± 0.09 | 1.70 ± 0.19 | 0.31 ± 0.05 |
| Years without descriptions (%) | 66 | 45 | 79 |
| Years with more than two species described (%) | 8 | 23 | 2 |
| Variance explained ($R^2 \times 100$) by the asymptotic model (%) | 98.72 | 98.74 | 98.51 |
| Total species predicted by the asymptote | 160 | 473 | 76 |
| Estimated percent of current described species | $\cong 100$ | 84 | 95 |
| Date to describe 50% of predicted species | 1838 | 1902 | 1860 |
| Date to describe 90% of predicted species | 1938 | 2066 | 1979 |

Results

How many species are yet to be discovered?

The number of species descriptions per year differs significantly between dung beetle families (Mann Whitney U -test, $Z = 2.71$, $P < 0.01$; $Z = 4.93$, $P < 0.001$; Z

= 7.42, $P < 0.001$ for Scarabaeidae–Geotrupidae, Scarabaeidae–Aphodiidae and Aphodiidae–Geotrupidae comparisons, respectively). The number of Aphodiidae species described per year (1.70) is greater than the number of Scarabaeidae (0.70), which in turn is greater than that of Geotrupidae (0.31). Within the period considered, no Aphodiidae species was described in half of those years, and in approximately one-quarter of the years, more than two species were described (Table 1). Only within one-third of the years considered was a new Scarabaeidae species described, while a new Geotrupidae species was described in only one-fifth of all the years. Seldom were more than two new species of these families described within the same year (Table 1, Figure 2).

The asymptotic fit of the relationships by the Beta–p function accounts for more than 98% of variance in the cumulative number of described species for the three families (Table 1, Figure 2). The total number of Geotrupidae species currently known is slightly lower than the number predicted by the asymptote. The opposite is the case with Scarabaeidae. However, for Aphodiidae, the known number of species is approximately 84% of the predicted total number of species. Half of the total predicted Scarabaeidae and Geotrupidae species had been described by the middle of the 19th century, while half of the total Aphodiidae species were described almost 50 years later (Table 1). For both Scarabaeidae and Geotrupidae, the rate of species description per year before 1850 (0.97 and 0.37, respectively) is higher than after 1850 (0.52 and 0.27), but the differences are not statistically significant for Geotrupidae (Mann Whitney U -test, $Z = 1.03$, $P = 0.30$) or Scarabaeidae ($Z = 1.40$, $P = 0.16$). To the contrary, the rate of Aphodiidae species description per year is significantly lower before 1850 (1.34) than after 1850 (1.93) ($Z = 2.21$, $P = 0.03$).

Modelling the description year

Both the backward selection procedure and the exhaustive search of multiple regression models for Scarabaeidae yield a function with GRS, maximum latitude and maximum longitude as significant variables. The Akaike information criterion score for this model (AIC = 1700.2) is the second lowest; 42.8% of total deviance is explained (Table 2). GRS is negatively and linearly related to the year of Scarabaeidae species description (Figure 3A). To the contrary, the negative relationships between the year of species description and maximum latitude and longitude are cubic and quadratic, respectively (Figure 3B and C). The cubic function of maximum latitude explains 18.7% of total deviance in the date of description (as compared with 16.1% for the linear function, Table 2), while the quadratic function of maximum longitude explains 22.1% of total deviance (as compared with 19.9% for the linear function). A final non-linear function model increases the total explained deviance only slightly (44.6 vs. 42.8%).

Both the backward selection procedure and exhaustive search agree on the significance of GRS, maximum latitude and maximum longitude variables in Aphodiidae (AIC = 4073.9; 56.9% of total deviance; Table 2). The relationships between year of description and BS and GRS are linear and negative (Figure 3D and

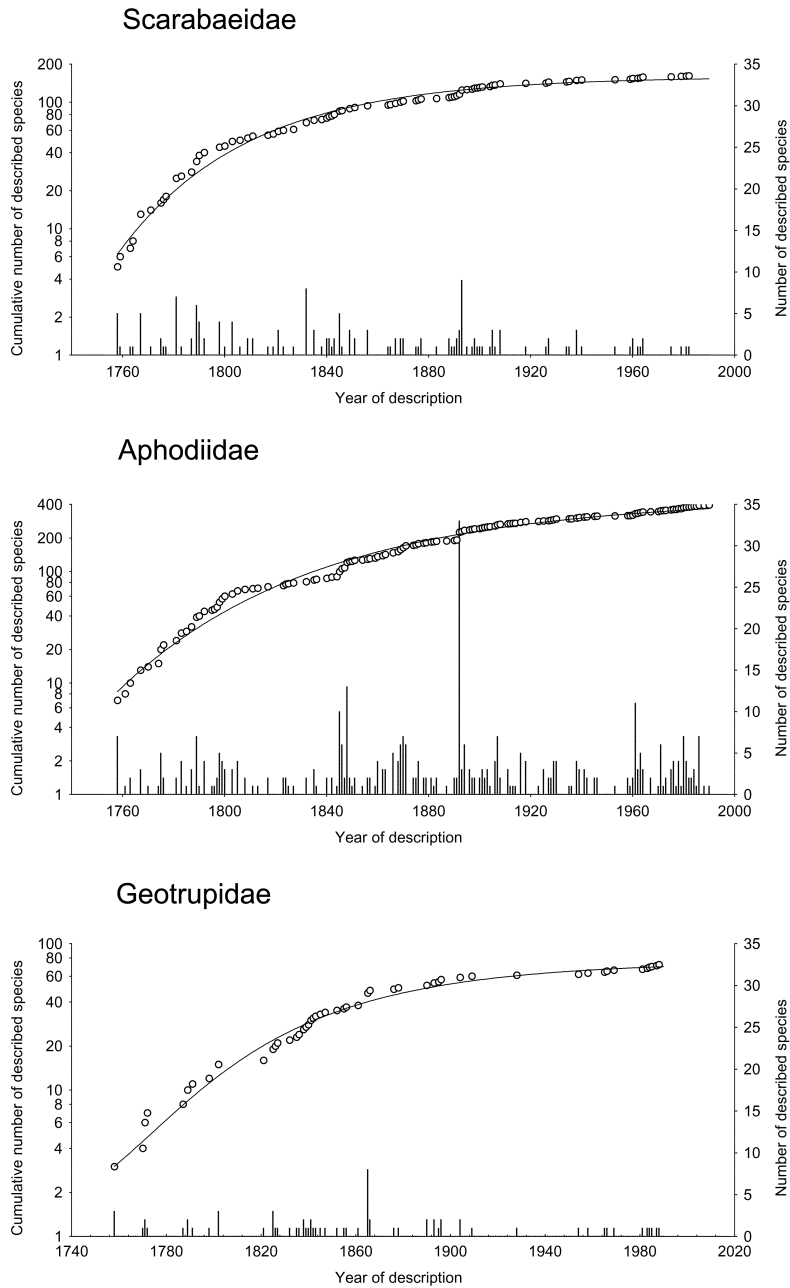


Figure 2. Number of described species (bars) and temporal variation in the logarithm of the cumulative number of described dung beetle species (○) from 1758 to 1990 in the Western Palearctic region, taking into account the exhaustive taxonomic information about the Palearctic species of the three Scarabaeoidea families: Scarabaeidae (162 species), Aphodiidae (395 species), and Geotrupidae (72 species). The cumulative curves were fitted (continuous line) by the Beta-p function (Mielke and Johnson 1974) using the Quasi-Newton method.

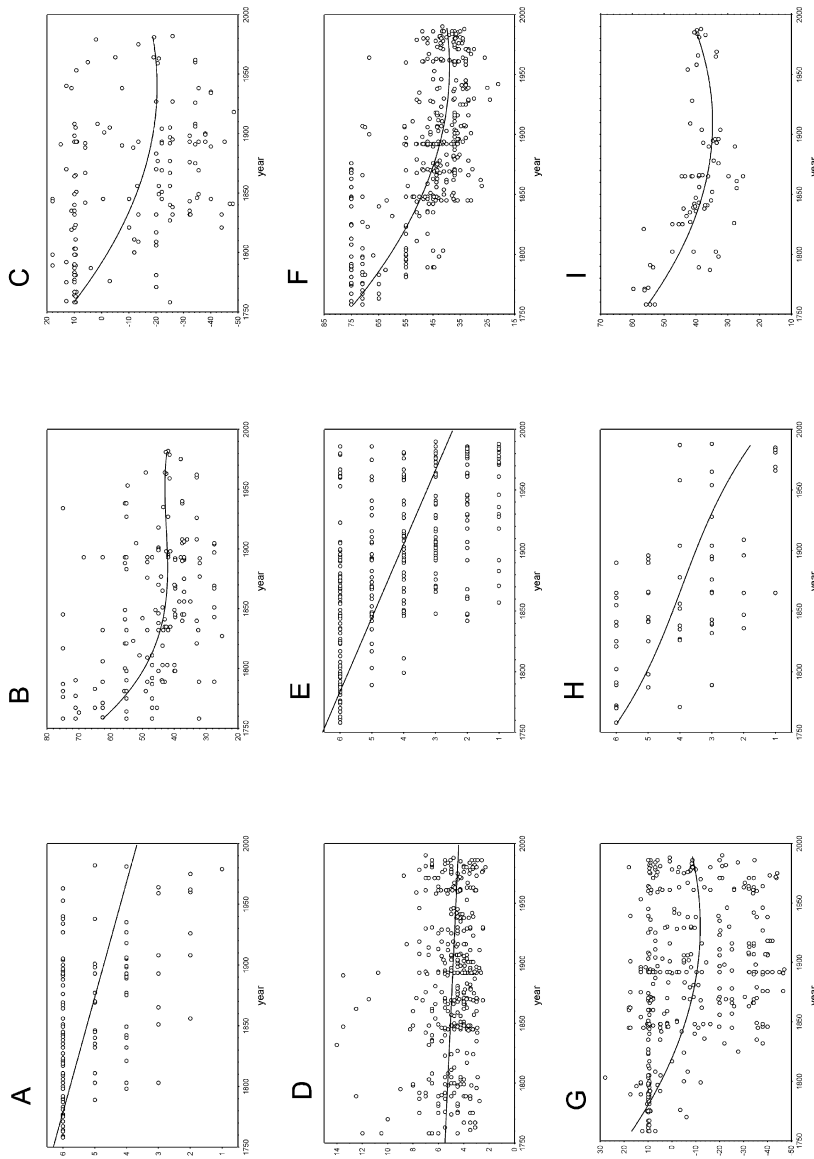


Figure 3. Scatterplots between the date and the statistically significant variables related to year of species description of the three dung beetle families considered (Table 2). (A–C) Statistically significant explanatory variables related to the year of Scarabaeidae species description. (D–G) Plots of the significant explanatory variables related to the year of Aphodiidae species description. (H, I) Plots of the two variables that significantly explain the variation in the year of Geotrupidae species description. The continuous line represents the selected linear, quadratic or cubic function of the variable considered that significantly increases the explanatory capacity of the model.

Table 2. Regression coefficients of statistically significant variables related to the year of species description of the three considered dung beetle families, and percentage of the total explained deviance (see Dobson 1999) when each variable is tested separately applying GLM regressions.

| | Parameter estimates | Standard error | Wald statistic ^a | P | Explained deviance (%) |
|-------------------|---------------------|----------------|-----------------------------|--------|------------------------|
| Scarabaeidae | | | | | |
| GRS | -18.90 | 3.23 | 34.27 | <0.001 | 27.3 |
| Maximum latitude | -0.97 | 0.32 | 8.89 | 0.003 | 16.1 |
| Maximum longitude | -1.02 | 0.18 | 30.68 | <0.001 | 19.9 |
| Aphodiidae | | | | | |
| BS | -4.83 | 1.23 | 15.43 | <0.001 | 2.5 |
| GRS | -14.07 | 1.56 | 81.77 | <0.001 | 39.0 |
| Maximum latitude | -1.90 | 0.20 | 91.83 | <0.001 | 43.2 |
| Maximum longitude | -0.58 | 0.11 | 26.15 | <0.001 | 10.9 |
| Geotrupidae | | | | | |
| GRS | -21.96 | 3.04 | 52.37 | <0.001 | 42.1 |
| Mean latitude | -2.99 | 0.65 | 21.43 | <0.001 | 23.0 |

^aThe Wald statistic is a test of the significance of the regression coefficient, based on maximum likelihood estimates.

E). However, maximum latitude and longitude are negatively and curvilinearly related to the year of species description (Figure 3F and G). The coefficients of the three terms of a cubic function of maximum latitude are significant; the explained deviance is increased from 43.2% (linear function) to 45.0%. The quadratic function of maximum longitude increases the total explained deviance from 10.9% (linear function) to 16.0%. The final curvilinear function model increases the explained deviance to 58.8%.

Exhaustive search and the backward stepwise procedures coincide in yielding the same Geotrupidae regression model (AIC = 746.1; 55.4% of total deviance) with GRS and mean latitude as significant variables (Table 2). The relationship between the year of Geotrupidae species description and GRS is linear and negative, while the relationship with mean latitude is negative but parabolic (Figure 3H and I). The quadratic function of mean latitude explains 33.6% of total deviance, and the final quadratic function model slightly increases the deviance explained to 55.6%.

Comparative importance of explanatory variables

The average effect of the incorporation of BS into all possible models is almost negligible (2.2, 1.8 and 0.9% of total deviance for Scarabaeidae, Aphodiidae and Geotrupidae, respectively). To the contrary, the contribution of GRS is decisive in explaining variation in the date of species description for the three families (17.6, 22.6 and 37.9%, respectively). The two main, significant GL variables (maximum or mean latitude and longitude) play different roles in the taxonomic groups. Maximum latitude explains comparatively less for Scarabaeidae (8.5%), but more for Aphodiidae (22.8% of total deviance). For Geotrupidae species, mean latitude is the GL variable that offers the most explanation (17.5%). Maximum longitude offers a

high level of explanation for Scarabaeidae (14.8%), but less for Aphodiidae (5.5%) and a negligible amount for Geotrupidae (0.5%).

The partitioning of variation among explanatory variables in Scarabaeidae (Figure 4A) corroborates that pure effects of GL and GRS variables account for a substantial part of the explained variation ($16.1\% + 9.9\% = 26\%$), and that the joint effect of these two types of variables explains 14.9% of the deviance. For Aphodiidae, the pure influence of GL and GRS is not very high (14.1% and 8.9%, respectively); the greater part of the explained variation is accounted for by the combined effect of both variables (31.4%; Figure 4B). The joint effect of GL and GRS is mainly due to the combined influence of latitude and GRS (Figure 5), because the most recently described species tend to have southern distribution limits and smaller range size areas (i.e. being a southern species and having a small distribution area are associated characteristics). Lastly, the pure effect of GL for Geotrupidae is similar to its effect for the other families (12.2%), but the pure effect of GRS is much greater (33.2%); the joint effect of GL and GRS is intermediate (10.1%; Figure 4C).

Discussion

The asymptotic fit of the cumulative number of species descriptions yields models that explain a high percentage of the variation of the date of species description (>98%). Although the total number of species predicted by the asymptote must be considered an approximation, the results indicate that the inventories of Western Palaearctic Scarabaeidae and Geotrupidae species are almost complete and that only a fraction of the species of these families have yet to be described. However, approximately 70–80 Aphodiidae species have not yet been discovered in the Western Palaearctic region. This implies that almost 16% of the total number of species comprising the Aphodiidae family have yet to be discovered.

As has been shown on other occasions (Allsopp 1997), the rate of species description is family-specific. The rate of species description in Aphodiidae has been more irregular (there is a greater variance in the number of species described per year), but also more constant over time (meanwhile, there are more years with new descriptions), as well as more rapid after 1850. Aphodiidae is the richest family in species among Palaearctic dung beetles (63% of total), possesses the smallest species' BS (approximate body length of 5 mm or less) and is the most difficult to identify. It is also significant that some species pertaining to the three studied families are not exclusively coprophagous in their adult stage (saprophagous or phytophagous habits are possible, for example) and Aphodiidae is precisely the family that comprises the larger group of these non-exclusive dung feeders, so these species could have been more tedious to locate. We suggest that taxonomic work probably would have turned to the Aphodiidae species once the description of new species in the other families had become more difficult.

For British beetles, body size offers more explanation (16%; Gaston 1991b). However, as in the Australian scarab beetles (Allsopp 1997), the species' BS

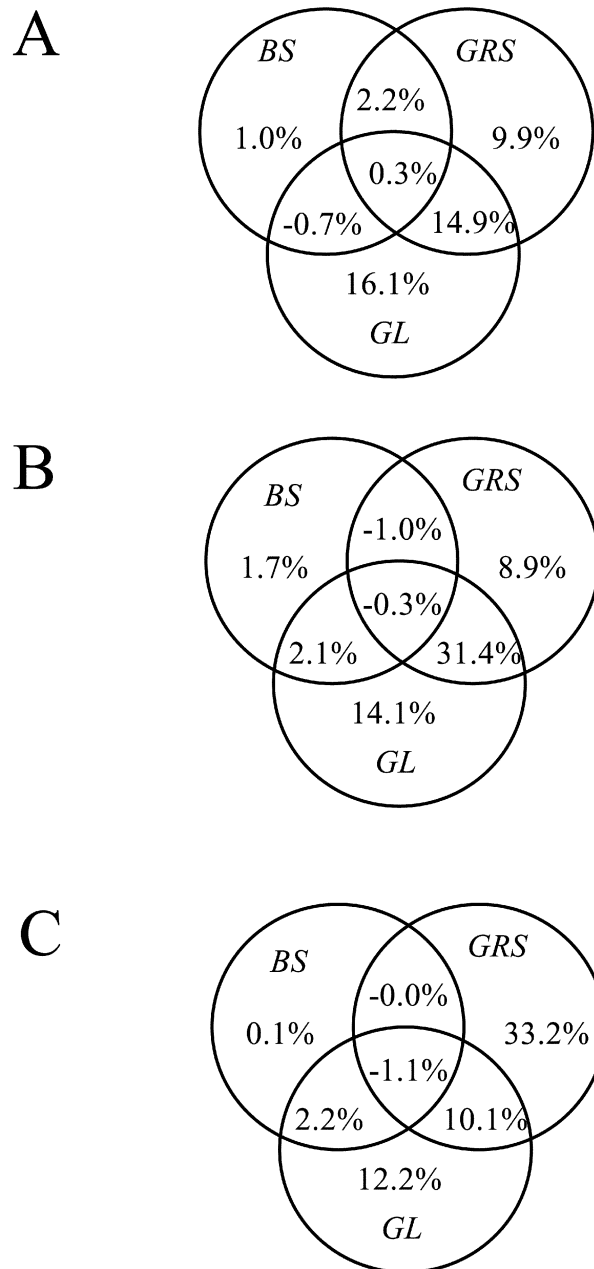


Figure 4. Variation partitioning in the year of species description of Scarabaeidae (A), Aphodiidae (B) and Geotrupidae (C) between the three considered explanatory variables: BS, GRS and GL. The decomposition of the variation has been carried out by partial regression analysis (Legendre and Legendre 1998; see Figure 1) using Generalized Linear Models (Crawley 1993) and the percentage of total explained deviance as the measure of explained variation.

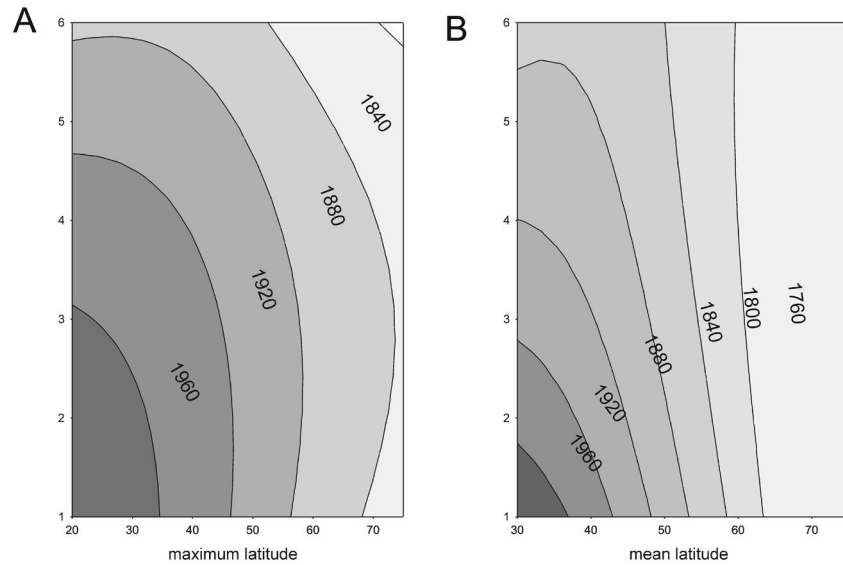


Figure 5. Contour plots of the year of species description for Aphodiidae (A) and Geotrupidae (B) in the space delimited by the scores of the GRS and maximum or medium latitude of their respective species. In both plots, it is clear that the most recently described species tend to have southern distribution limits and smaller range size areas.

accounted for a negligible fraction in the variation of the year of description for any of the three families (<2%). Our result might be an outcome of the family-specific analysis of the data, but also the consequence of the greater chance of discovery of new Scarabaeoidea species because they normally live in restricted habitats where they are usually sought (e.g. cattle excrement).

To the contrary, the GRS and the GL are essential variables negatively correlated with the year of species description, which together can account for high percentages in the year of description variation (from 41 to 56%). GRS has traditionally been considered an important factor linearly related to year of description, as it has been confirmed that broad-range species in a variety of groups are somewhat prone to being discovered before species with narrower ranges; for example, South American oscine passerine birds (Blackburn and Gaston 1995), North American butterflies (Gaston et al. 1995a), Neotropical mammals (Patterson 1994, 2000), or Australian scarab beetles (Allsopp 1997). GL has also been considered an important factor related to the probability of species description (Gaston 1993; Patterson 1994; Gaston et al. 1995b; Allsopp 1997). Its relationship with the date of our studied species description is curvilinear at all times. Species with distributions reaching the northern and western limits of the Palearctic region are described first, but from approximately 1850 on, the distribution limits of the greater part of the species described are between 30°–50° latitude and 10°–30° eastern longitude.

However, geographical variables influence the families differently. For Scarabaeidae, the GRS and the GL variables combined, as well as the pure effect of

range size primarily determine the year of species description. The pure effect of GL variables is also important; both maximum distributional range, longitude and latitude strongly influence the year in which species were described. For Aphodiidae species, GRS combined with GL (principally, species maximum latitude) primarily determine the description of new species; wide geographical range species also capable of inhabiting high latitudes and eastern longitudes are the first to be discovered. For Geotrupidae, the pure effect of range size undoubtedly best predicts species description.

Why have variables such as range size, maximum latitude or maximum longitude so decisively influenced the description of new species in the Western Palaearctic region? As is the case with many other groups, both Nearctic and Western Palaearctic dung beetle assemblages decrease significantly in species richness with latitude (Hanski 1986, 1991; Lobo 2000); dung beetle species with narrow geographical ranges are more abundant in Southern Palaearctic areas (Lumaret and Lobo 1996). These regions, refuges during glaciations, would have supplied colonizers for northern areas (Bennett et al. 1991; Hewitt 1996; Taberlet et al. 1998). Thus, only some generalist species, with broader distribution ranges at present, would have managed to colonize northern areas of the West Palaearctic region. Taxonomists are not homogeneously distributed across regions either (Gaston and May 1992; Gaston 1994). Western and northern countries in the West Palaearctic region have, throughout history, been home to established taxonomy. Hence, taxonomists often first described northwestern and broader-ranged generalist species, because the more a species' geographical distribution overlaps the taxonomists' location, the greater its chance for early description.

In summary, the search for new Western Palaearctic dung beetle species must be directed towards the Aphodiidae family. Extrapolation of accumulation curves (which should be treated with caution) indicates that if description continues at its present rate, it would take as long as 400 years to discover the total number of Aphodiidae species. The current information could identify the features of species not yet discovered and where to find them. As geographical distribution and location combined are so important for this family, new species should be found to the south and east in the Western Palaearctic region in very small areas. The best areas to search for new Western Palaearctic dung beetle species should be in north and northeast Africa, southeast Europe, Turkey, Caucasus, Caspian regions, and Iran. These possibly new, not-yet-described species could be easterly species shared with the Eastern Palaearctic region. The geographical distribution of Western Palaearctic region species described since 1991 (Zoological Record database) corroborates these patterns. During the 1991–2000 period, a total of 13 new species were described; six Aphodiidae (in north Africa, eastern Europe, Turkey and Near East, and Iran) and seven Geotrupidae species (see Figure 6). Three of these new Geotrupidae species come from the Iberian Peninsula, but were obtained when newer taxonomic tools and technological advances were applied to previously described species; three more from the Trans-Caspian region, shared with the Eastern Palaearctic fauna; and one from north Africa.

Finally, variation partitioning is proposed as a useful procedure to study the

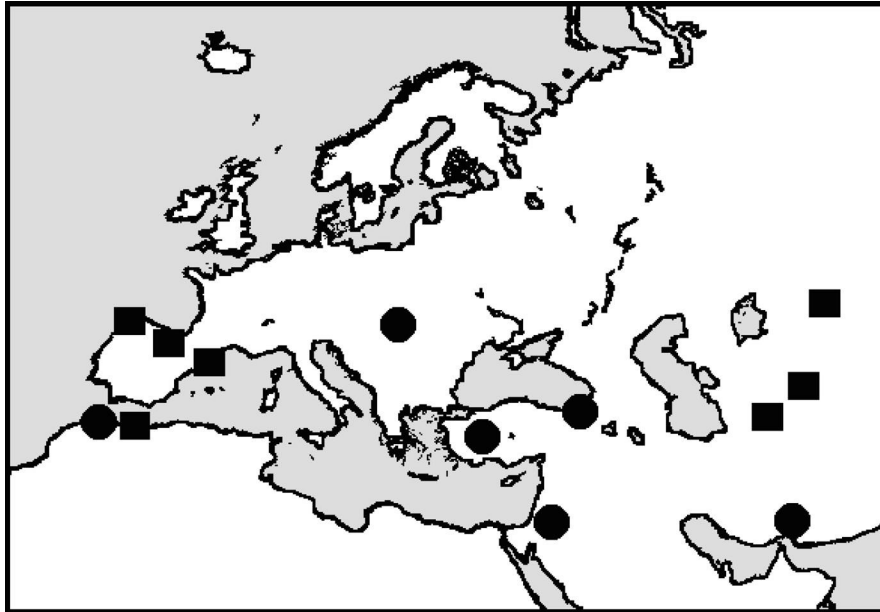


Figure 6. Western Palearctic map showing type localities of newly described species from 1990 to the present. Geotrupidae: ■; Aphodiidae: ●.

contribution of different variables that may influence the rate of species description for any taxonomic group, allowing us to discern the pure effects of the combined contribution of these variables so that future inventory efforts could be planned more efficiently.

Acknowledgements

This paper was supported by the Spanish D.G.I. Grant REN2001-1136/GLO and a Ph.D grant from the Ministerio de Educación y Cultura y Deporte (F.J.C.-S.) associated with the PB97-1149 project.

References

- Akaike H. 1983. Information measures and model selection. *Bulletin of the International Statistical Institute: Proceedings of the 44th Session 1*: 277–290.
- Allsopp P.G. 1997. Probability of describing an Australian scarab beetle: influence of body size and distribution. *Journal of Biogeography* 24: 717–724.
- Anderson M.J. and Gribble N.A. 1998. Partitioning the variation among spatial, temporal and environmental components in a multivariate data set. *Australian Journal of Ecology* 23: 158–167.
- Arnett R.H. 1967. Present and future systematics of the Coleoptera in North America. *Annals of the Entomological Society of America* 60: 162–170.

- Arnone M., Carpaneto G.M. and Piattella E. 1995. Arthropoda di Lampedusa, Linosa e Pantelleria (Canale di Sicilia, Mar Mediterraneo). Coleoptera Scarabaeoidea. *Naturalista Siciliana* 19: 447–468.
- Arnone M. and Massa B. 1993. Note biologiche e corologiche su alcuni *Aphodius* Illiger in Sicilia. *Naturalista Siciliana* S. IV 17: 271–289.
- Baraud J. 1985. Coléoptères Scarabaeoidea. Faune du Nord de l'Afrique du Maroc au Sinai. Editions Lechevalier, Paris.
- Baraud J. 1992. Coléoptères Scarabaeoidea d'Europe. Société Linnéenne de Lyon, Lyon, France.
- Barrowclough G.F. 1992. Systematics, biodiversity, and conservation biology. In: Eldregde N. (ed.), *Systematics, Ecology and the Biodiversity Crisis*. Columbia University Press, New York, pp. 1–14.
- Bennett K.D., Tzedakis P.C. and Willis K.J. 1991. Quaternary refugia of north European trees. *Journal of Biogeography* 18: 103–115.
- Blackburn T.M. and 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.
- Borcard D., Legendre P. and Drapeau P. 1992. Partialling out the spatial component of ecological variation. *Ecology* 73: 1045–1055.
- Carpaneto G.M. 1976. Reperti di Coleotteri Scarabaeoidea Laparosticti del vicino oriente e delle Isole Greche. *Fragmenta Entomologica* 12: 255–271.
- Carpaneto G.M. 1979. Osservazioni sui coleotteri Scarabaeoidea delle Isole Ponziane. *Fragmenta Entomologica* 15: 111–126.
- Carpaneto G.M. 1980. I Coleotteri Scarabeoidei endemici del complesso sardo-corso. *Lavori della Società Italiana di Biogeografia* 8: 675–690.
- Carpaneto G.M. 1985. Dati faunistici e considerazioni zoogeografiche sugli Scarabeoidei delle Isole Eolie (Coleoptera, Scarabaeoidea). *Animalia* 12: 87–99.
- Carpaneto G.M. and Piattella E. 1995a. Osservazioni zoogeografiche sui Coleotteri Scarabaeoidei delle piccole Isole Circumsarde (Coleoptera, Scarabaeoidea). *Biogeographia* 18: 441–454.
- Carpaneto G.M. and Piattella E. 1995b. Coleoptera Polyphaga V (Lucanoidea, Scarabaeoidea). Checklist delle specie della fauna italiana. Edizioni Calderini, Bologna, Italy.
- Carpaneto G.M., Dellacasa G., Dellacasa M., Piattella E. and Poggi R. 1997. Ricerche zoologiche della nave oceanografica 'Minerva' (C.N.R.) sulle Isole Circumsarde. XXVI. Catalogo faunistico e zoogeografico dei coleotteri Scarabaeoidei (Coleoptera, Scarabaeoidea). *Annali del Museo Civico di Storia Naturale 'G. Doria'* 91: 511–539.
- Chikatunov V. and Pavlíček T. 1997. Catalogue of the beetles (Coleoptera) in Israel and adjacent areas: 1. *Klapalekiana* 33: 37–65.
- Copas J.B. 1983. Regression, prediction and shrinkage (with discussion). *Journal of the Royal Statistical Society B* 45: 311–354.
- Cotterill F.P.D. 1995. Systematics, biological knowledge and environmental conservation. *Biodiversity and Conservation* 4: 183–205.
- Crawley M.J. 1993. *GLIM for Ecologists*. Blackwell Scientific Publications, Oxford, UK.
- Daniel C. and Wood F.S. 1980. *Fitting Equations to Data*. John Wiley & Sons, New York.
- Dellacasa G. 1990. The genera *Apsteiniella*, *Cnemargulus*, *Cnemisus* and their type species (Scarabaeoidea Aphodiidae). *Frustula Entomologica* 13: 19–30.
- Dellacasa G. 1992. Systematic redefinition of *Aphodius* subgenera *Mendidius* Harold, *Sitiphus* Fairmaire and *Sugrames* Reitter (Coleoptera Aphodiidae). *Frustula Entomologica* 15: 119–132.
- Dellacasa G. and Poggi R. 1981. Materiali per una fauna dell'Arcipelago Toscano. XXV. Elenco degli Scarabaeoidea dell'Isola del Giglio (Coleoptera). *Annali del Museo Civico di Storia Naturale di Genova* 83: 443–454.
- Dellacasa M. 1987. Aphodiini delle Isole Eolie. *Naturalista Siciliana* S. IV 11: 1–4.
- Dellacasa M. 1988a. Contribution to a world-wide catalogue of Aegialiidae, Aphodiidae, Aulonocnemidae, Termitotrogidae (Coleoptera Scarabaeoidea). *Memorie della Società Italiana* 66: 1–455.
- Dellacasa M. 1988b. Contribution to a world-wide catalogue of Aegialiidae, Aphodiidae, Aulonocnemidae, Termitotrogidae (Coleoptera Scarabaeoidea). Part II. *Memorie della Società Italiana* 67: 1–231.
- Dellacasa M. 1988c. Contribution to a world-wide catalogue of Aegialiidae, Aphodiidae, Aulonoc-

- nemidae, Termitotrogidae (Coleoptera Scarabaeoidea). *Addenda et Corrigenda (First Note)*. *Memorie della Società Italiana* 67: 291–316.
- Dellacasa M. 1991. Contribution to a world-wide catalogue of Aegialiidae, Aphodiidae, Aulonocnemidae, Termitotrogidae (Coleoptera Scarabaeoidea). *Addenda et Corrigenda (Second Note)*. *Memorie della Società Italiana* 70: 3–57.
- Dellacasa M. 1995. Contribution to a world-wide catalogue of Aegialiidae, Aphodiidae, Aulonocnemidae, Termitotrogidae (Coleoptera Scarabaeoidea). *Addenda et Corrigenda (Third Note)*. *Memorie della Società Italiana* 74: 159–232.
- Derksen S. and Keselman H.J. 1992. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology* 45: 265–282.
- Diamond J.M. 1985. How many unknown species are yet to be discovered? *Nature* 315: 538–539.
- Dobson A. 1999. *An Introduction to Generalized Linear Models*. Chapman & Hall/CRC, London.
- Flather C.H. 1996. Fitting species-accumulation functions and assessing regional land use impacts on avian diversity. *Journal of Biogeography* 23: 155–168.
- Frank J.H. and Curtis G.A. 1979. Trend lines and the number of species of Staphylinidae. *The Coleopterists Bulletin* 33: 133–168.
- Gaston K.J. 1991a. The magnitude of global insect species richness. *Conservation Biology* 5: 283–296.
- Gaston K.J. 1991b. Body size and probability of description: the beetle fauna of Britain. *Ecological Entomology* 16: 505–508.
- Gaston K.J. 1993. Spatial patterns in the description and richness of the Hymenoptera. In: Lasalle J. and Gould I.D. (eds), *Hymenoptera Biodiversity*. C.A.B. International, Wallingford, UK.
- Gaston K.J. 1994. Spatial patterns of species description: how is our knowledge of the global insect fauna growing? *Biological Conservation* 67: 37–40.
- Gaston K.J. 1996. Species richness: measure and measurement. In: Gaston K.J. (ed.), *Biodiversity. A Biology of Numbers and Difference*. Blackwell Science, Oxford, UK.
- Gaston K.J. and Blackburn T.M. 1994. Are newly described bird species small-bodied? *Biodiversity Letters* 2: 16–20.
- Gaston K.J. and May R.M. 1992. Taxonomy of taxonomists. *Nature* 356: 281–282.
- Gaston K.J. and Mound L.A. 1993. Taxonomy, hypothesis testing and the biodiversity crisis. *Proceedings of the Royal Society of London B* 251: 139–142.
- Gaston K.J., Blackburn T.M. and Loder N. 1995a. Which species are described first? The case of North American butterflies. *Biodiversity and Conservation* 4: 119–127.
- Gaston K.J., Scoble M.J. and Crock A. 1995b. Patterns in species description: a case study using the Geometridae (Lepidoptera). *Biological Journal of the Linnean Society* 55: 225–237.
- Hammond P. 1992. Species inventory. In: Groombridge B. (ed.), *Global Biodiversity*. World Conservation Monitoring Centre. Chapman & Hall, London.
- Hanski I. 1986. Individual behaviour, population dynamics and community structure of *Aphodius* (Scarabaeidae) in Europe. *Acta Oecologica, Oecologia Generalis* 7: 171–187.
- Hanski I. 1991. North temperate dung beetles. In: Hanski I. and Cambefort Y. (eds), *Dung Beetle Ecology*. Princeton University Press, Princeton, New Jersey, pp. 75–96.
- Hanski I. and Kuusela S. 1983. Dung beetle communities in the Aland archipelago. *Acta Entomologica Fennica* 42: 36–42.
- Heard S.B. and Mooers A.Ø. 2000. Phylogenetically patterned speciation rates and extinction risks change the loss of evolutionary history during extinctions. *Proceedings of the Royal Society of London B* 267: 613–620.
- Hewitt G.M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58: 247–276.
- Heywood V.H. (ed.) 1995. *Global Biodiversity Assessment*. Cambridge University Press, Cambridge, UK.
- Hollande A. and Théron J. 1998. *Aphodiidae du Nord de l'Afrique (Coleoptera Scarabaeoidea)*. Monograph XXI. Museo Regionale di Scienze Naturali, Torino, Italy.
- Hughes J.B., Daily G.C. and Ehrlich P.R. 1997. Population diversity: its extent and extinction. *Science* 278: 689–692.

- Illies J. 1983. Changing concepts in Biogeography. *Annual Review of Entomology* 28: 391–406.
- Král D. 1997. A review of Chinese *Aphodius* species. Part 4: subgenera *Pseudacrossus* and *Qingaphodius* sbg. n. (Coleoptera: Scarabaeidae). *Acta Societatis Zoologicae Bohemicae* 61: 129–149.
- Král D. and Olexa A. 1996. New and otherwise noteworthy *Lethrus* species from Central Asia and Afghanistan (Coleoptera: Geotrupidae). *Folia Heyrovskyana* 4: 49–65.
- Krell F.-T. 2000. Impact factors are not relevant to taxonomy. *Nature* 405: 507–508.
- Lawton J.H. and May R.M. 1995. *Extinction Rates*. Oxford University Press, Oxford, UK.
- Legendre P. and Legendre L. 1998. *Numerical Ecology*. 2nd edn. Elsevier, Amsterdam.
- Lobo J.M. 1993. Estimation of dung beetle biomass (Coleoptera, Scarabaeoidea). *European Journal of Entomology* 90: 235–238.
- Lobo J.M. 2000. Species diversity and composition of dung beetle (Coleoptera: Scarabaeoidea) assemblages in North America. *The Canadian Entomologist* 132: 307–321.
- Lobo J.M. and Martín-Piera F. 1993. Análisis comparado de las comunidades primaverales de escarabeidos coprófagos (Col., Scarabaeoidea) del archipiélago balear. *Ecología Mediterránea* 19: 29–41.
- López-Colón J.I. 1995. El 'Género' *Thorectes* Mulsant, 1842 (Coleoptera, Scarabaeoidea, Geotrupidae) en la Fauna Europea. *Giornale Italiano di Entomologia* 7: 355–388.
- Lumaret J.P. 1990. Atlas des Coleopteres Scarabeides Laparosticti de France. *Inventaires de Faune et de Flore* 1: 1–418.
- Lumaret J.P. and Kirk A.A. 1987. Ecology of dung beetles in the French Mediterranean region (Coleoptera, Scarabaeidae). *Acta Zoologica Mexicana (N.S.)* 24: 1–60.
- Lumaret J.P. and Lobo J.M. 1996. Geographic distribution of endemic dung beetles (Coleoptera, Scarabaeoidea) in the western Palaearctic region. *Biodiversity Letters* 3: 192–199.
- Lumaret J.P., Lobo J.M. and Jay-Robert P. 1996. Catalogue and atlas of the Scarabaeoidea Laparosticti endemic to France. *Collection Patrimoines Naturels* 26: 1–70.
- Mackauer M. 1958. Eine coprophagen-ausbeute aus dem noerdlichen sizilien (Coleoptera Scarabaeidae). *Memorie della Società Entomologica Italiana* 37: 46–54.
- MacNally R. 2000. Regression and model-building in conservation biology, biogeography and ecology: The distinction between – and reconciliation of – 'predictive' and 'explanatory' models. *Biodiversity and Conservation* 9: 655–671.
- Martín-Piera F. and Lobo J.M. 1992. Los Scarabaeoidea laparosticti del archipiélago Balear (Coleoptera). *Nouvelle Revue d'Entomologie* 9: 15–28.
- May R.M. 1988. How many species are there on Earth? *Science* 241: 1441–1449.
- Mayr E. 1993. The inventory of natural diversity. In: Minelli A. (ed.), *Biological Systematics. The State of the Art*. Chapman & Hall, London.
- McCullagh P. and Nelder J.A. 1989. *Generalized Linear Models*. 2nd edn. Chapman & Hall, London.
- Medellín R.A. and Soberón J. 1999. Predictions of mammal diversity on four land masses. *Conservation Biology* 13: 143–149.
- Mielke P.W. and Johnson E.S. 1974. Some generalized beta distributions of the second kind having desirable application features in hydrology and meteorology. *Water Resource Research* 10: 223–226.
- Neter J., Wasserman W. and Kutner M.H. 1985. *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*. Irwin, Homewood, Illinois.
- Nicholls A.O. 1989. How to make biological surveys go further with generalised linear models. *Biological Conservation* 50: 51–75.
- O'Brien C.W. and Wibmer G.J. 1979. The use of trend curves of rates of species descriptions: examples from the Curculionidae (Coleoptera). *The Coleopterists Bulletin* 33: 151–166.
- Palmer M., García Ple C. and Morey M. 1988. Distribución estacional de las especies del género *Aphodius* (Coleoptera, Scarabaeoidea) en las heces de ovino de Mallorca. I: Abundancia. *Reunión Científica de la S.E.E.P. (Sociedad Española para el Estudio de los Pastos) XXVIII*: 155–167.
- Patterson B.D. 1994. Accumulating knowledge on the dimensions of biodiversity: systematic perspectives on Neotropical mammals. *Biodiversity Letters* 2: 79–86.
- Patterson B.D. 2000. Patterns and trends in the discovery of new Neotropical mammals. *Diversity and Distribution* 6: 145–151.

- Pittino R. 1996. An interesting rediscovery: *Osmanius balthasari* (Coleoptera, Scarabaeoidea, Aphodiidae). *Fragmenta entomologica* 27: 355–360.
- Purvis A., Aapow P.M., Gittleman J.L. and Mace G.M. 2000. Nonrandom extinction risk and the loss of evolutionary history. *Science* 288: 328–330.
- Qinghong L. and Bråkenhielm S. 1995. A statistical approach to decompose ecological variation. *Water, Air and Soil Pollution* 85: 1587–1592.
- Skidmore P. 1991. *Insects of the British Cow-Dung Community*. Field Studies Council Publications, Shrewsbury, UK.
- Soberón J. and Llorente J. 1993. The use of species accumulation functions for the prediction of species richness. *Conservation Biology* 7: 480–488.
- Scholtz C.M. 1990. Phylogenetic trends in the Scarabaeoidea (Coleoptera). *Journal of Natural History* 24: 1027–1066.
- StatSoft Inc. 1999. *STATISTICA. Computer Program Manual for Windows*. StatSoft Inc., Tulsa, Oklahoma.
- Steyskal G.C. 1965. Trend curves of the rate of species description in zoology. *Science* 149: 880–882.
- Sutton S.L. and Collins N.M. 1991. Insects and tropical forest conservation. In: Collins N.M. and Thomas J.A. (eds), *The Conservation of Insects and their Habitats*. Academic Press, London, pp. 405–424.
- Taberlet P., Fumagalli L., Wust-Saucy A.G. and Cosson J.F. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7: 453–464.
- White R.E. 1975. Trend curves of the rate of species description for certain North American Coleoptera. *The Coleopterists Bulletin* 29: 281–295.
- White R.E. 1979. Response to the use of trend curves by Erwin, Frank and Curtis, and O'Brien and Wibmer. *The Coleopterists Bulletin* 33: 167–168.
- Ziani S. 1997a. Scarabaeoidea Laparosticti di Cipro: un primo quadro d'insieme sulla fauna dell'isola (Coleoptera). *Biocosme Mésogéen* 14: 1–38.
- Ziani S. 1997b. Novità e precisazioni su alcuni Scarabaeoidea del bacino del Mediterraneo. *Bollettino della Società Entomologica Italiana*, Genova 128: 197–200.