From DNA barcodes to ecology: meta-analysis of central European beetles reveal link with species ecology but also to data pattern and gaps

Sara Ottati[1], Jonas Eberle[2], Frank Köhler[3], Björn Rulik[1], and Dirk Ahrens[1]

[1]Zoologisches Forschungsinstitut und Museum Alexander Koenig
[2]Paris Lodron University of Salzburg
[3]n/a

September 13, 2022

Abstract

DNA barcoding has been used worldwide to identify biological specimens and to delimit species. It represents a cost-effective, fast and efficient way to assess biodiversity with help of the public Barcode of Life Database (BOLD) accounting for more than 236,000 animal species and more than ten million barcode sequences. Here, we performed a meta-analysis of available barcode data of central European Coleoptera to detect intraspecific genetic patterns among ecological groups in relation to geographic distance with the aim to investigate a possible link between infraspecific variation and species ecology. We collected information regarding feeding style, body size as well as habitat and biotope preferences. Mantel tests and two variants of Procrustes analysis, both involving the Principal Coordinates Neighborhood Matrices (PCNM) approach, were applied on genetic and geographic distance matrices. However, significance levels were too low to further use the outcome for further trait investigation: these were in mean for all ecological guilds only 7.5, 9.4, or 15.6 % for PCNM+PCA, NMDS+PCA, and Mantel test, respectively, or at best 28% for a single guild. Our study confirmed that certain ecological traits were associated with higher species diversity and foster stronger genetic differentiation. Results suggest that increased numbers of species, sampling localities, and specimens for a chosen area of interest may give new insights to explore barcode data and species ecology for the scope of conservation on a larger scale.

Introduction

The maintenance of biodiversity, as the sum of all "plants, animals, fungi, and microorganisms on Earth, their genotypic and phenotypic variation, and the communities and ecosystems of which they are a part" (Dirzo & Raven, 2003), is one of the most important current concerns of humankind, as wild species are decreasing at an alarming rate, and an inversion of this trend requires an anthropic intervention to guarantee their survival (Frankham et al., 2002). Biodiversity is composed by multiple dimensions, and no single measure of biodiversity can capture all its dimensions (Carpenter et al., 2009): Genetic diversity is essential in order to develop an evolutionary potential for species to be able to react to environmental changes (Toro & Cabarello, 2005). However, very little is known on trends in genetic diversity, particularly in wild species (Pereira et al., 2012). While taxonomic coverage with indicator taxa and diversity assessments is very limited, the extinction risk of the vast majority of biodiversity is not known (Pereira et al., 2012). Thus, a characterization and management of genetic diversity seems necessary considering idiosyncratic population structures, as well as to choose the correct way and the proper resolution power to estimate it (Storfen et al., 2010).

Mitochondrial DNA (mtDNA) has been a marker of choice for reconstructing historical patterns of population demography, admixture, biogeography, and speciation (Castro et al., 1998; Hull & Jiggins, 2005). Mitochondrial DNA (mtDNA) is maternally inherited and generally a non-recombinant marker. Variation in

mtDNA is assessable by DNA sequencing in a cost-effective way which are exactly the properties that make mtDNA marker suitable for the large-scale assessment of species boundaries through its variation being used in rapid assessment approaches for biodiversity research (Ratnasingham & Hebert, 2007), such as barcoding studies. Barcoding uses a single gene fragment (*COI*) and, thus, through large scale barcoding universal data of intraspecific genetic variation became available over a wide range of organisms.

Ecological studies typically require a determination of the species involved. Acquisition of such biodiversity data for plants and animals using morphological characteristics to identify field collected samples requires both a significant time effort for identification based on morphology and sufficient taxonomic expertise that is rarely available for a vast variety of organism groups. Therefore, many ecological studies lack taxonomic information while ecological data for many species are rare. The recent development of DNA-based methods for species identification, known as DNA barcoding (Hebert et al., 2003a), has drastically simplified this identification step (Coissac et al., 2012) and might help to overcome the gap between taxonomy and ecology.

Using a standardized genetic marker in DNA barcoding allows connecting the identities of different life stages such as eggs, larvae, or adults – often a major difficulty in morphology-based taxonomy and cryptic species (e.g., Ahrens et al., 2007; Šipek & Ahrens, 2011; García-Robledo et al., 2013; Etzler et al., 2014; Köhler et al., 2022) and to trace in the environment remnants of organismal DNA (Taberlet et al., 2012; Yu et al., 2012). Barcoding has been successfully applied to a vast number of taxa in many different geographic regions (www.boldsystems.org). It has become obvious that validated, comprehensive species libraries are the most fundamental basis for optimal barcode-based taxon identification (Kvist, 2013). The huge amount of barcode data with large number of already collected, identified and barcoded specimens opens up to ecologists to use these data of biodiversity in a vast geographic scale. They enhance a fast and clear overview on the biodiversity. Indeed, population genetic analysis of ecological communities with *COI* sequences extends the value of the DNA barcode employed as identification and taxonomic tool. Whereas barcoding for taxonomic purposes was in past often limited by economic constraints to a very few individuals per species, larger comprehensive studies becoming more and more available (e.g., Dincă et al., 2011; Bergsten et al., 2012; Hendrich et al., 2015; Rulik et al., 2017). These data-rich studies with multiple sampling sites may provide useful population genetic information with a link to the entire species distribution range applicable to a range of ecological and historical questions (Craft et al., 2010; Baselga et al., 2013, 2015).

DNA metabarcoding, which couples the principles of DNA barcoding with next generation sequencing technology, provides an opportunity to easily produce large amounts of data on biodiversity. Microbiologists have long used metabarcoding approaches, but use of this technique in the assessment of biodiversity in plant and animal communities is under-explored. DNA metabarcoding, which couples the principles of DNA barcoding with next generation sequencing technology, provides an opportunity to easily produce large amounts of data on biodiversity. Microbiologists have long used metabarcoding approaches, but use of this technique in the assessment of biodiversity in plant and animal communities is underexplored.

In this study, we analyze random samples from different Central European beetle metapopulations in order to provide a first meta-analysis of data generated in a large-scale barcoding project (GBOL; Hendrich et al., 2015; Rulik et al., 2017) with focus on their intraspecific degree of genetic divergence compared to their spatial and ecological properties. We collected all the available metadata information about Central European beetles contained in the BOLD data base in order to provide a rapid, and as complete as possible, survey about the beetle biodiversity to explore patterns of genetic differentiation among different ecological guilds considering the spatial scale to investigate circumstances driving the intraspecific mtDNA divergence in beetle species (Coleoptera). We were interested whether the relation of intraspecific genetic distances and geographic distances differed between ecological guilds. Finally, we aimed to provide a new and complete all-fauna phylogeographical overview on middle European beetle biodiversity.

Material and methods

Data, species taxonomy and ecological information

The specimen's data for this study have been collected from two previous barcoding studies on central Euro-

pean beetles (Insecta: Coleoptera) (Hendrich et al., 2016; Rulik et al., 2017) performed in the framework of German Barcoding initiatives (i.e., German Barcode of Life project: https://www.bolgermany.de; Bavarian Barcoding project: http://www.faunabavarica.de/). These projects aim at building a reference library of DNA barcodes for all available organisms in Germany collecting, where possible, ten specimens per species, from locations as distinct as possible throughout the Germany and the neighboring countries in order to capture genetic variability (Figure 1).

In order to avoid an underestimation of intraspecific genetic differences and to maximize the amount of complete available data from BOLD, (http://www.barcodinglife.org), we excluded from the analysis redundant genetic information represented by identical syntopic haplotypes with equal geographical coordinates and incomplete ecological or geographical information on the specimens. From these, we retained only those species for which were available comprehensive ecological information (see below).

Species taxonomy used as backbone for this study is derived from Klausnitzer & Köhler (1998) and subsequent works (Köhler, 2000, 2011a, b; Bleich et al., 2016) reflecting the current species taxonomy applied in German coleopterist's community (http://www.coleokat.de/de/fhl/ (status: 2016)). Eventual inconsistencies of current classification with the source of ecological data (Koch 1989, 1991, 1992) were adopted by F.K. in his curated data base with help of the numerical identifier for each species (Lohse & Lucht, 1999). For our meta-analysis, we selected among available data to examine the effect of four major ecological variables (body size, biotype preference, habitat preference, and feeding habits) in the context of geographical distance and genetic (mtDNA) differentiation.

Mitochondrial DNA (mtDNA) has been widely used in phylogenetic studies of animals because it evolves much more rapidly than most nuclear DNA, resulting in the accumulation of differences between closely related species (Brown et al., 1979; Moore, 1995). The rapid pace of sequence evolution in mtDNA, and in particular in *COI* , results in differences between populations that have only been separated for brief periods of time, making it a relevant way to infer divergence at population level (Avise et al., 1987). *COI* has been found to consistently differentiate species and for which large libraries of sequences in constant growth have become available linked to voucher specimens (Hebert et al., 2003a, b; Hebert & Gregory, 2005).

For the autecological information on the species (Table S1), which we regard here as a proxy for the entire species ecology, we used data on feeding habits, habitat preference, biotope preferences, and body size (derived from a data base curated by F.K. based mainly on Koch, 1989, 1991, 1992 and many further, more detailed publications not mentioned here). Feeding habits included eight different, more generic classes: coprophagous, polyphagous, mycetophagous, necrophagous, phytophagous, saprophagous, xylophagous, and zoophagous (Figure 3B). While the same number of classes was used for the habitat preference: soil; eurytop, rotting matters, nests, mushrooms, vegetation, water, and dead wood (Figure 3A), the biotope preferences were represented by four different categories: no biotope preference, wetlands, open-land biotypes, and forests (Figure 3C). Five size classes were arbitrary defined: 0-2 mm, 2-5 mm, 5-10 mm, 10-20 mm, 20-50 mm. These reflect roughly eco-functional groupings in the food web due to body size. However, therefore, the ecological classes are not equally represented among the sampled species (Figure 3D).

Metadata analysis

In order to inquire which ecological species properties influence the intraspecific genetic differentiation, we investigated for correlation between infraspecific genetic distances of *COI* barcode data and geographical distances among infraspecific samples using three different approaches (Figure 2, see below). Due some controversy in the scientific community about Mantel test (Peres-Neto et al., 2001; Legendre et al., 2015), we included here also ordination techniques (Legendre & Legendre 2012) and combined these to the method of principal coordinates of neighbor matrices in a Procrustes analysis (PCNM; Borcard & Legendre, 2002; Borcard et al., 2004; Dray et al., 2006). In a second step, we linked the outcome of these comparisons to the ecological properties for each species. Significance is expected to be in major part is also dependent from intraspecific sampling of each species which is in many species rather limited. The sheer amount of available data nevertheless promises sufficient significant results.

Meta-analysis was performed using R (R Core Team, 2017) in R-studio (Rstudio Team, 2016), Euclidean distance matrices from the set of geographical coordinates of collection sites of each specimen were generated with *geosphere* package (Hijmans et al., 2016). Pairwise distances from *COI* sequences using the Kimura 2-parameters DNA substitution model (Kimura, 1980) were calculated in the *Ape*package (Paradis et al., 2014). Even there was some controversy about the use of Kimura 2-parameters as DNA substitution model in barcodes (Magnacca & Brown, 2010; Moniz & Kaczmarska, 2010; Srivathsan & Meier, 2012), we use this substitution model as recommended by Hebert et al. (2003) since this study inquires at "population" level.

On these two distance matrices we subsequently performed the following analyses to infer the number of species of each ecological class whose intraspecific genetic distances were correlated with the geographical distances: 1) a Mantel test, two variants of Procrustes test using either 2) Principal coordinates of neighbor matrices (PCNM) and non-metric multidimensional scaling (NMDS) or, 3) transformation-based principal components analysis (Tb-PCA) and NMDS. PCNM represents the geographical distance matrix while PcoA and NMDS represent the genetic distance matrix. Statistical significance in these comparisons indicates strength of evidence about the probable trend of differentiation of the populations involved (Allendorf & Luikart, 2007).

The Mantel test is a multivariate statistic method which typically compares two distance matrices that were calculated for the same set of objects but that are based on two independent sets of variables (e.g., a species dissimilarity matrix and site distance matrix) (Mantel, 1967). The test calculates the correlation between values in the corresponding positions of two matrices. Significance of the linear relationship between matrices is assessed through permutation of objects (Peres-Neto, 2001). Being first applied in population genetics by Sokal (1979), the Mantel test is currently one of the most commonly used methods to evaluate the relationship between geographic distance and genetic divergence (Mantel, 1967; see Manly, 1985, 1997; Diniz-Fhilo et al., 2013) – despite recent controversy and criticism about its statistical performance (e.g., Harmon & Glor, 2010; Legendre & Fortin, 2010; Guillot & Rousset, 2013, Catellano & Balletto, 2002) and the existence of more sophisticated and complex approaches to analyze spatial multivariate data (Diniz-Fhilo et al., 2013) – the Mantel test is still one of the most employed method in matrix data correlations. While Oden & Sokal (1992) reported a problem for the partial Mantel test for spatially autocorrelated data, Lapointe (1995) found problems with the simple Mantel test when using it for the comparison of dendrograms, and Peres-Neto et al. (2001) tested error type I on Mantel statistic compared to other techniques. Lastly, Nunn et al. (2006) and Harmon & Glor (2010) expressed concerns about the simple and partial Mantel tests when used for phylogenetic comparative analyses. Essentially, all the issues reported by these studies relate to inflated type I error rate or low power (Guillot & Rousset, 2013). In regard of the study design, the quality and the amount of the analyzed data of our case study, the Mantel test seems to be a reasonable technique- to inquire on the degree of relation between the two distance matrices. We ran the Mantel test on submatrices, each one representing a single species composed from more than four specimens. The null hypothesis, i.e., the absence of relationship, was rejected when the p value was lower than 0.05. To run these analyses we used the function "mantel()" in the vegan package (Oksanen et al., 2017) where the Mantel statistics was defined as a matrix correlation between the two, geographical and genetic, dissimilarity matrices.

Ordination techniques provide several alternative ways to search for a correlation among different kind of matrices (Legendre & Legendre, 2012). The most common current alternative to the Mantel tests is to ordinate the genetic distances and compare them with geographic coordinates or other vector representations of geographical distances (Diniz-Fhilo et al., 2013). Principal coordinates of neighbor matrices (PCNM; Borcard & Legendre, 2002; Borcard et al., 2004; Dray et al., 2006), is a powerful approach able to detect the spatial structure of varying scale in response data. Essentially, spatial variables are used to determine the distance between sites with special focus on neighboring sites (Borcard & Legendre, 2002). These distances are decomposed into a new set of independent (and hence orthogonal) spatial variables; it can be considered as a more general approach to transform geographic space in a raw data form. There are several variants of this approach (see Griffith & Peres-Neto, 2006; Bini et al., 2009; Landeiro et al., 2011; Diniz-Filho et al., 2009, 2012). The methodology followed here is the spatial Eigenfunction analysis (SEA) which has been extensively used in ecology, and also gained attention from landscape geneticists (i.e., Manel et al., 2003; Manel &

Holderegger, 2013). PCNM can detect a wide range of spatial structures, including autocorrelation and periodic structures (Dray et al., 2006). The distance between objects is represented as a Euclidean distance matrix, calculated from spatial data (latitude and longitude values) associated with the sample locations. As the name suggests, PCNM is primarily concerned with 'neighboring' sites. We used the PCNM() tool in the vegan package, that automatically set the threshold distance above which distances are simply considered "large". Any geographical distances above this value were set to four times the threshold value (see Borcard & Legendre, 2002). Finally, the modified distance matrix was subject to principal coordinates analysis (PcoA). Due to the 'truncation' of the original distance matrix to create a neighbor matrix, a PcoA on a neighbor matrix (typically) produce more eigenvectors relative to the same analysis on a standard distance matrix. All resulting eigenvectors with positive eigenvalues have been used as a new set of explanatory spatial variables in a multiple regression approach trough Procrustes superimposition method. Therefore, when genetic data are regressed against these eigenvectors, some of them tend to describe the spatial patterns in genetic variation (Legendre & Fortin, 2010).

The transformation-based principal components analysis (Tb-PCA) is a two-step ordination method that it is considered to be the two-step equivalent of the principal coordinates analysis (PcoA). The principal coordinates analysis (PcoA) is a conceptual extension of the principal components analysis (PCA) technique (Pearson, 1901) also integrated in the PCNM procedure (Figure 2). It similarly seeks to order the objects along the axes of principal coordinates while attempting to explain the variance in the original data set. However, while PCA organizes objects by an eigenanalysis of a correlation or covariance matrix, PcoA can be applied to any distance (dissimilarity) matrix (Gower, 1966). As the PCNM technique, PCA generates a set of eigenvectors summarizing the distance matrix information. Legendre & Gallagher (2001) recommended, in order to simplify the analysis and standardize the data entries, to detrend the data (decostand() function in "vegan" package) before performing the PCA. Only the resulting positive eigenvectors have been combined with PCNM eigenvectors in a Procrustes analysis, which aims to find the species where the two distance matrices are better correlated.

Non-metric multidimensional scaling (NMDS) is a unique ordination technique in that a (small) number of ordination axes are explicitly chosen prior to the analysis and the data are then fitted to those dimensions. Because NMDS is a numerical rather than an analytical technique, it does not produce a unique solution. A 'stress' parameter is computed to measure the lack of fit between object distances in the NMDS ordination space and the calculated dissimilarities among objects (Paliy & Shankar, 2016). In contrast to tb-PCA, NMDS is a non-eigenvector-based approach, i.e., all axes produced are equally representing the data variance (Legendre et al., 2005). In order to compare and verify the reliability of the results we have chosen to perform this ordination method as well as the eigenvector-based one, and to compare the results then with each other. We performed NMDS analysis aiming at transforming the genetic distance matrices in a set of row data representing the genetic distance variance for every species (Figure 2).

In order to investigate the impact of sampling area size on the results, a fifth categorical, sampling-dependent variable have been introduced: the distance classes which represent the average distance between all individuals of the same species. A high distance class is equivalent with a larger sampling area of a species. The distance class thus represents the distance in which most of the sampled specimens were found. The largest geographical distance value was found to be 719 km (*Longitarsus parvulus* Paykull, 1799), the smallest was about 5 km (*Chrysolina marginata* Linnaeus, 1758). All mean distances within 100 km were in the distance class "100", all those within 101 and 150 km were in the distance class "150", those between 151 and 200 km in class "200", to class "300" belong all those species ranging between 201 and 300 km and the last class includes all the species with a mean geographic distance higher than 301 km.

At first, we performed Procrustes analysis on PCNM axes and tb-PCA axes, representing the geographical structure and the genetic distances of each species, respectively (the latter calculated on the detrended data frame as recommended by Borcard & Legendre, 2002) (Figure 2). We have chosen Procrustes superimposition technique because variables could be either ordination axes (usually those that explain most of the variation in a data set) or original variables; it may also be applied to (dis)similarity matrices describing the same objects.

Procrustes analysis is a statistical method which compares a collection of (multidimensional) shapes by attempting to transform them into a state of maximal superimposition. It does so by attempting to minimize the sum of squared distances between corresponding points in each shape through translation, reflection, rigid rotation, and dilation (scaling) of their coordinate matrices. In contrast to the Mantel test, Procrustes analysis allows to determine how much variance in one matrix is attributable to the variance in the other. When comparing ordinations from NMDS or PCA a 'shape' may be defined by treating each ordinated point as a vertex (Peres-Neto & Jackson, 2001). In the present context, Procrustes analysis is applied on matrices that are back-converted into an "object-by-variable" table by principal coordinates analysis (PcoA), non-metric multidimensional scaling (NMDS), or other suitable ordination methods (Jackson 2001). Configurations resulting from Procrustes analysis can be tested for non-randomness through repeated symmetric Procrustes analyses allowing to test for significant differences (protest()-function in vegan). We repeated the same analysis using the NMDS axes instead of the tb-PCA eigenvectors (both representing the genetic distance).

Finally, we performed linear regressions one by one within the same species on the sets of axes representing genetic distances against the PCNM axes representing the geographic distances to test for significance of the relationship between the two variables. PCNM axes were computed both times using pcnm() function in vegan package (Oksanen et al., 2017), while tb-PCA and NMDS axes were computed respectively using the pcoa() function in vegan and the nmds() function in "ecodist" package (Goslee & Urban, 2007).

Regarding the results of Procrustes analysis, we refer in the following text (statistics and significance values) to "NMDS statistics", "NMDS significance" or generally to "NMDS results" for results that were computed using regressing NMDS and PCNM axes, while we use "PCA statistics", "PCA significance" and "PCA results" for results referring to the Procrustes output involving tb-PCA axes and PCNM eigenvectors.

Role of ecological characteristics

Both pairwise intraspecific geographical and genetic distances were plotted in subsets of ecological properties (see above). Subsequently, sampling variables (i.e., the number of sampled localities, the number of sampled specimens and the mean geographic distance among sampled individuals per species) were plotted against test statistics and significance to investigate the degree of their correlation. To check for the role of sampling design on the meta-analyses through plot choice, we compared statistics and significance scores of every species against the other variables with the number of individuals examined per species, the number of sampled localities and the mean geographic distance among sampled individuals per species.

To detect potential population genetic "anomalies" or "regularities" referable to ecological dynamics, we produced boxplots on all data and on subsets of data, observing the variance in the mean values of statistic and significance scores obtained as Mantel/NMDS/PCA output. We were aware that significance in major part is also dependent from the amount of intraspecific sampling of each species and the strength of the relation.

Results

Of the 29,464 available *COI* barcode sequences, geographical coordinates were available for 29,349 individuals (Figure 1). Excluding identical syntopic haplotypes and individuals with incomplete geographical coordinates, we retained for the analysis 16283 individuals of 3967 species from 124 Coleoptera families. From these, we retained only those species for which comprehensive ecological information was available, such as feeding style, habitat and biotope preference, and body size resulting in 12207 specimens available for final meta-analysis representing 1785 species and 95 families of Coleoptera from Germany.

The majority of the species (above 400) was distributed in the distance class "300", the fewest species were in the class ">300" (Figure S1). The number of species decreased exponentially with the rising of the number of localities, with one exception where only 190 species recorded 3 sampling localities (Figure S1). Most species were zoophagous (n= 674) or phytophagous (n= 585), while the preferred habitat was vegetation, which made up more than 600 species (n= 632) while dead wood and soil habitats accounted for 398 and 400 species, respectively (Figure S2). Across the preferred biotopes, 691 species preferred forest, many others

either open-land (n= 420) or wetland (n=369), and 305 had not specific preferences (Figure S2). The most represented size-classes were, as expected, very small (n= 594) and small (n= 626) species. The medium size class accounted for 386 species, while the large and largest beetles counted 157 and 22 species, respectively (Figure S2).

Infraspecific genetic distances resulted to be inversely proportional to the number of sampled individuals and number of sampling localities (Figure 3A-C), while intraspecific geographic distances were not affected by sampling issues. Despite numerous cases with elevated infraspecific divergence, distance plots across the four major eco-classes of central European beetles (habitat, feeding preferences, biotope, size class) revealed generally low intraspecific distances (<3 % sequence divergence). Nevertheless, at first glance, patterns of geographical and genetic differentiation differ clearly among ecological and eco-morphological traits. Indeed, the plots (all data) of geographical distance matrices versus genetic matrices showed different shapes across different ecological guilds (Figure 3B-H, Figure S3), particularly for habitat and feeding style. However, for the biotope types no principal differences were visually evident (Figure 3G), although in biotopes, wetland and open land included more long-distance samples than did forest and eurytopic biotopes. Body size showed a clear trend that with the increase of body size (from extra-small to extra-large) followed a mostly gradual decrease of the infraspecific genetic distances (Figure 3F, Figure S3C). Larger geographical distances occurred in saprophagous and phytophagous taxa or species preferring wetland and vegetation, for which specimens had been sequenced from outside of Germany (Figure S3). In saprophagous taxa, a large amount of pairwise distances of specimens followed a nearly proportional increase in both genetic and geographic distances. The same was found for the xylophagous and zoophagous species. A different pattern was evident in mycetophagous species where the genetic distances were already much higher even with short geographic distances, while in necrophagous, polyphagous and coprophagous species we observed the opposite (with some minor exceptions), i.e., a clearly limited amount of genetic variation even with increased geographic distance (Figure 3D). In regard of habitat preferences, vegetation, rotten matters, and dead woods had similar trends as taxa with phyto-, sapro-, and xylophagous feeding style, obviously due to the connection between these eco-classes. Species with habitats such as soil and nests had also a higher genetic variation at same geographical scale, while others such as the "mushrooms/ fungi" inhabitants had comparatively lower genetic structure despite a vast geographical sampling (Figure 3E).

All the three methods of correlation analysis between genetic and geographical differentiation detected cases in which the expected relationship between geographical and genetic distances were confirmed, i.e., where all correlation approaches were positive (Figures S4, S5). The Mantel test identified 250 species having a significant relationship between the two distances (14% of the total number of examined species), NMDS 160 species (9% of the total number of examined species) and PCA 123 (6.8% of the total number of examined species). Overlap of significant species between NMDS and Mantels test was 9.2%, while species overlap of PCA with Mantel test was 9.6 % of and 32.5 % with NMDS. This apparent discordance in identifying the significant species with correlated patterns was evident from pairwise plots of significance and statistic values of each species from the three methods (Figure S6): while the PCA and NMDS methods agreed in both significance and statistic values in most of the cases, both methods highly diverged with the mantel resulting scores. Furthermore, maximum and minimal values of significant species differed between the three methods (Tab. 1).

Significance for the ecological traits and their sub-guilds was rather rare in all three approaches of correlation analysis (Figure S5). Significance levels were too low to further use the outcome for further trait investigation: these were in mean for all ecological guilds 7.5, 9.4, or 15.6 % for PCNM+PCA, NMDS+PCA, and Mantel test, respectively, or at best 28% for a single guilt (Tab. 2). Some beetle families entirely did not show significance (Figure S7). All three correlation tests found coprophagous species to have the significantly more species with significant correlation between genetic and geographic distances (Tabs 1, 2). While the mantel test identified nest dwellers with the lowest percentage the opposite was the case for NMDS and PCNM where the nest dweller guild was the most dominant in significance. PCA and NMDS recorded high percentages in significant species in nest habitat and body size (i.e., extra-large), while small species were those dominant for the Mantel test.

Discussion

Patterns of genetic variation often reflect spatial variation in gene flow which can be influenced in two important ways (Wang & Summers, 2013): Spatially separated populations may experience isolation-by-distance (IBD; Wright 1943) in which landscape barriers and geographical distances cause restricted gene flow; and isolation-by-environment (IBE; Wang & Summers, 2010) in which gene flow among populations inhabiting different ecological environments is limited either by selection against dispersers moving between them or by individual preference to remain in a particular environment due to local adaptation (Dobzhansky, 1937). IBE predicts a correlation between genetic divergence and environmental dissimilarity because greater environmental differences between populations are expected to be associated with stronger divergent selection and reduction in the success of dispersers (Crispo et al., 2006; Lee & Mitchell-Olds, 2011). Of course, geographical and environmental isolation do not exclude each other, and spatial genetic divergence can be associated to gene flow reduction due to both geographical and ecological factors (e.g., Coyne & Orr, 2004; Crispo et al., 2006; Thorpe et al., 2008). Furthermore, population genetic theory predicts that genetic distances among individuals will increase with increasing geographical distance (Allendorf & Luikart, 2007).

Our investigation of the genetic vs. geographic distances within species across multiple guilds of different ecological traits and taxa (95 Coleoptera families) revealed that there are only few cases in which this relationship resulted to be significant and thus predictable, i.e., for which an increase in geographic intraspecific distances was followed by an increase in genetic distances, or the other way around, and most of all, showing a sufficient sampling (in terms of number of samples to examine) to support such trends statistically. However, observed patterns of infraspecific genetic and geographical distances among most of the central European Coleoptera species and ecological guilds examined were neither uniform nor entirely different among each other. Thus, the unlikely hypothesis that all species increase their genetic diversity with the distance of their record, which could be interpreted as a signal of gradual dispersion and genetic differentiation in progress (Allendorf & Luikart, 2007), could be not universally confirmed, despite the study area suffered an almost entire biodiversity loss during the Pleistocene and was reoccupied afterwards from external founder populations (e.g., Hewitt, 2000; Hofreiter & Stewart, 2009; Abellán et al. 2011; Birks & Tinner, 2016).

Besides limited sampling, cases, in which species did not show a positive relation between genetic and geographical distances, might be explained in two ways: 1) the geographic expansion of the species was not followed by an equal genetic diversification (intraspecific DNA distances are smaller than geographic distances), i.e., occurring in species with high dispersal capabilities with continuous genetic mixing and in a well interconnected area. Or, 2) the genetic diversification in the study area was independent from the geographic scale, occurring mainly in species with a potentially different phylogeographic origin of their populations. It is well known that Central Europe experienced post-glacial recolonization events from different Mediterranean and extra-Mediterranean refugia located all across the continent (e.g., Ahrens et al., 2013; Kühne et al., 2017, Schmitt & Varga, 2012). Freijeiro & Baselga (2016) suggested that dispersal-based processes in European beetles were probably taxon-dependent, but also depended on dispersal ability and ecological traits (Gómez-Rodríguez et al., 2015). Although patterns appear not very clear due to widely lacking significance, we here found several patterns in genetic/geographic relationships among ecological preference classes which might fit more ecology-dependent dispersal and differentiation (Papadopoulou et al., 2008). Indeed, causalities are expected to depend both on environmental and ecological processes in the species range. The distribution area as well as the relation genetic distance vs. geographic distance of a species depends on several factors: the paleo-biogeographic and biogeographic history (i.e., glacial expansion dynamics, glacial refugia presence, postglacial climatic gradients, and postglacial species expansion) (Stewart et al., 2010) could have been the major cause of such genetic trend in diversification. It is widely accepted that present distribution patterns in Central Europe are related to post glacial recolonization dynamics (e.g., Schuldt & Assmann, 2009; Schmitt, 2009) beside other also important factors (Baselga et al., 2012). In this context, geographic, climatic and ecological exogenous factors (i.e., climatic gradients, habitat fragmentation or presence/absence of corridors) and ecological endogenous factors (i.e., potential niche or dispersal capabilities due to physiological properties, level of adaptation) play a crucial role to determine these patterns (e.g., Schmitt et al., 2009; Rundell et al., 2009). So far, distribution patterns and genetic differentiation have been studied for mainly

selected cases in the framework of phylogeographic studies (taxa or/and study sites; e.g., Múrria et al., 2020; Garcia-Raventós et al., 2021; Domènech et al., 2022), while only some studies with wider taxonomic and geographical scope exist (e.g., Baselga et al., 2013, 2015; Joly et al., 2014; Fujisawa et al., 2014; Dapporto et al., 2019). With the upcoming barcode data, a vast amount of data is becoming available to address such questions routinely at large scale, and to uncover particularly responses at population level regarding many ecological and climatic factors which have so far been explored with limited systematic sampling.

Here, deeper going conclusions lack statistic support since barcode data/ libraries are generally not designed to explore phylogeographic patterns in the context of species ecology. At this stage we expect sampling bias since data were generated with the scope of collecting and barcoding as many species as possible for future species identification. However, the amount of available data on central European beetle species was good enough to start to enquire the relationships between ecological properties of the species and their intraspecific patterns of genetic differentiation and to look at patterns that go beyond a single guild or species group (Baselga et al., 2013, 2015; Fujisawa et al., 2014). In fact, we faced severe problems due to the available number of sampling localities and of individuals per species. Therefore, we included all the sampling variables in the linear models excluding all the species with poor number of specimens and sampling sites. Furthermore, we investigated for the role of the sampling variables, such as number of individuals per species and number of localities, on the final results of statistic and significance scores as well as for an eventual implication of geographic distances of arbitrary chosen sample sites.

PCA and NMDS techniques captured different information compared to the Mantel test. Thus, the PCA technique was the least efficient in describing the relation of genetic and geographic distances in terms of amount of resulting significant species, followed by NMDS method and Mantel test. Results of both were similar but partially different from those obtained from Mantel test (Figure S5). This discordance limited more general conclusions. It is likely that the different algorithms behind the analytical techniques behaved differently in the presence of high level of noise in the data caused by spatial autocorrelation (Diniz-Filho et al., 2003; Legendre et al., 2015; although not tested here) and lacking sufficient geographical sampling (due to financial constraint of the Barcoding initiative, which did not allow higher samples numbers). Because different methods may emphasize different aspects of the data, using different data analyses techniques (Figure 2) may reveal more aspects of the data structure than a single method (Kenkel & Orloci, 1986). The Mantel test was considered to handle the limited number of available samples per species best which was disadvantageous for the ordination technique, which better read and converted the data matrices in more readable and efficient row data (Legendre et al., 2015) for the further Procrustes analysis. Ordination techniques are known to work better when dealing with big amount of data. On the other hand, minimal sampling size in our data was below the generally suggested amount to robustly investigate phenomena depending on spatial scale (at least 20 sampling localities; Dale & Fortin, 2014). Being first applied in population genetics by Sokal (1979), the Mantel test is currently one of the most commonly used methods to evaluate the relationship between geographic distance and genetic divergence (Mantel, 1967; see Manly, 1985, 1997; Diniz-Fhilo et al., 2013) – despite recent controversy and criticism about its statistical performance (e.g., Harmon & Glor, 2010; Legendre & Fortin, 2010; Guillot & Rousset, 2013; Castellano & Balletto, 2002) and the existence of more sophisticated and complex approaches to analyze spatial multivariate data (Diniz-Fhilo et al., 2013). In our case study, the mean number was only five sampling localities per species. Even though ordination methods are better suited, less prone to type I error and better in describing patterns (Legendre & Fortin, 2010; Legendre et al., 2015; C. Wang et al., 2010; I.J. Wang et al., 2013), results were not congruent with those of the Mantel tests. Nevertheless, PCNM methods combined to genetic information should be considered an alternative to the Mantel test and further analysis on a richer dataset could then possibly lead to clearer ecological conclusions.

It is known that the occupied habitat type has significant effects on both extent of the species range and latitudinal distribution (Ribera & Vogler, 2000; Hof et al., 2006, Fujisawa et al., 2014). This extends by some aspects the results of Fujisawa et al (2014) who found infraspecific genetic variation of *COI* in water beetles positively correlated with occupancy (numbers of sites of species presence; i.e., a similar but not identical measure to geographical distance) and negatively with latitude, whereas substitution rates across species

(which we did not examine here) was influenced mainly by habitat types; specialized species of more stable environments, such as running water, had the highest rate. Baselga et al. (2015) expected dispersal to be high in aquatic beetles (of standing waters) because of the need for movement between ephemeral water bodies, while dispersal of leaf beetles do not require long-range movement for population persistence due to more stable conditions in vegetation. This is also reflected by our findings for species using vegetation as habitat. Our data thus seem to confirm the habitat stability hypothesis (Ribera et al., 2003) which sees in Pleistocene glacial events and the following climatic stability the major causes in producing equilibrium conditions, either with environmental factors due to niche-based processes or with spatial distributions from long-term stochastic dispersal.

Our data suggested higher dispersal tendencies and lower infraspecific variation of mtDNA for more ephemeral food resources (dung, dead animals), or habitats (fungi/ mushrooms). However, low number of species in these guilds and a similar pattern for eurytopic species (Tab. 1) might indicate that this observed pattern could be also a result of sampling bias. Specimens' body size does not provide an answer to this question, as generally divergent patterns of infraspecific genetic vs geographical distances between smaller (x_s, s, m) and larger species (l, x_l) (Figure 3) are contrasted by rather uniform correlation statistics between the size classes (Figure 4). Studies on ground beetles have shown a generally higher genetic diversity across larger species independent from their sample site distance (Assmann et al., 2010; Schuldt & Assmann, 2011) which explain this pattern by lack of interconnection among populations due to their very specific habitat requirements, the habitat quality, and respective morphological adaptations (e.g., wing reduction; Jelaska & Durbešić, 2009). According to Freijeiro & Baselga (2016), the presence or absence of wings is an important factor for a better understanding of the geographical/genetic scale relationship. Indeed, habitat fragmentation is considered a major factor limiting gene flow in ground beetle populations (Liebherr, 1988).

Our results, beside yet still enormous sampling gaps and underrepresentation of many species, indicate that ecological niches and preferences may play a major role in geographic dispersal and genetic differentiation within species even though we did not consider here environmental, climatic factors or longitudinal\latitudinal gradients which are also known to have a fundamental role in explaining population dynamics (Rosindell et al., 2011; Baselga et al., 2013, 2015; Frejeiro & Baselga, 2016). These results can be extremely helpful to further develop conservation strategies, from a simple species conservation approach towards the conservation of genetic diversity in habitats or landscapes (e.g., Hedrick, 2001; Vellend et al., 2014). Therefore, further molecular screening would be needed, with particular focus on more geographical sampling to cover more in detail the genetic variation within the study area and to uncover causalities of such patterns (i.e., extending the Barcoding towards population level). Indeed, our results showed: an increase in number of sampling localities was usually followed by a related increase in statistic score and thus increase the explanatory power of barcode data to explain infraspecific genetic patterns among different ecological guilds.

The screening at the diversity patterns of the entire entomological fauna in such vast territory as Central Europe is a demanding task which request efficiency and great sampling efforts, but in the light of the emergency of current trends of insect decline (e.g., Hallmann et al., 2017; Wagner et al., 2021) it becomes an important issue for deeper understanding of its causes. The German Barcoding of Life campaign (Hendrich et al., 2015; Rulik et al., 2017) and resulting database contributed to resolve these issues. Nevertheless, a denser geographic sampling that will also result from future monitoring studies or metabarcoding projects will enhance the number of sampled localities and specimens, while concerted actions would be desirable. This will strengthen statistic results and allow bolder conclusions regarding biodiversity in a study area rather than simple species numbers.

Acknowledgements

This project was funded by the German Federal Ministry of Education and Research (work package GBOL1; FKZ: 01LI1101A / 01LI1501A).

References

Abellán, P., Benetti, C.J., Angus, R.B., Ribera, I. (2011). A review of Quaternary range shifts in Euro-

pean aquatic Coleoptera. *Global Ecology and Biogeography* , 20, 87–100. https://doi.org/10.1111/j.1466-8238.2010.00572.x

Ahrens, D., Monaghan, M.T., Vogler, A.P. (2007). DNA-based taxonomy for associating adults and larvae in multi-species assemblages of chafers (Coleoptera: Scarabaeidae). *Molecular Phylogenetics and Evolution,* 44, 436–449.

Ahrens, D., Fabrizi, S., Sipek, P., Lago, P. (2013). Integrative analysis of DNA phylogeography and morphology of the European rose chafer (*Cetonia aurata*) to infer species taxonomy and patterns of postglacial colonisation in Europe. *Molecular Phylogenetics and Evolution* , 69, 83–94.

Allendorf, F.W., Luikart, G. (2007). *Conservation and the Genetics of Populations* . Malden (MA): Blackwell.

Assmann, T., Casale, A., Drees, C., Habel, J., Matern, A., Schuldt, A. (2010). The dark side of relict species biology: cave animals as ancient lineages. In: Habel J, Assmann T. (Eds) *Relict species: Phylogeography and conservation biology* . Springer, Berlin, Heidelberg, 91–103.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A., Saunders, N.C. (1987). Intraspecific phylogeography: The mitochondrial bridge between population genetics and systematics.*Annual Review of Ecology, Evolution, and Systematics* , 18, 489–522.

Baselga, A., Lobo, J.M., Svenning, J.-C. and Araújo, M.B. (2012). Global patterns in the shape of species geographical ranges reveal range determinants. *Journal of Biogeography* , 39, 760–771. https://doi.org/10.1111/j.1365-2699.2011.02612.x

Baselga, A., Fujisawa, T., Crampton-Platt, A., Bergsten, J., Foster, P.G., Monaghan, M.T., Vogler, A.P. (2013). Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels. *Nature Communications* , 4, 1892.

Baselga, A., Gómez-Rodriguez, C., Vogler, A.P. (2015). Multi-hierarchical macroecology at species and genetic levels to discern neutral and non-neutral processes. *Global Ecology and Biogeography* , 24, 873–882.

Bergsten, J., Bilton, D.T., Fujisawa, T., Elliott, M., Monaghan, M.T., Balke, M., Ribera, I., Nilsson, A.N., Barraclough, T.G., Vogler, A.P. (2012). The effect of geographical scale of sampling on DNA barcoding.*Systematic Biology* , 61, 851–869.

Bini, M.L., Diniz-Filho, J.A.F., Rangel, et al. (2009). Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. *Ecography* , 32, 193-204.

Birks, H. J. B., Tinner, W. (2016). Past forests of Europe. In: San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, A. (Eds.) *European atlas of forest tree species* . Publ. Off. EU, Luxembourg, pp. E010c45+

Bleich, O., Gurlich, S., Kohler, F. (2016). *Verzeichnis der Kafer Deutschlands* . *Entomofauna Germanica.*http://www.coleokat.de/de/fhl/.

Borcard, D., Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices.*Ecological Modelling* , 153, 57–68.

Borcard, D., Legendre, P., Carol, A.-J., Hanna, T. (2004). Dissecting the spatial structure of ecological data at multiple scales.*Ecology* , 85, 1826-1832.

Brown, W.M., George, M. Jr., Wilson, A.C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences USA* , 76, 1967–1971. Doi:10.1073/pnas.76.4.1967

Carpenter, S.R., Mooney, H.A., Agard, J., Capistrano, D., Defries, R.S., Diaz, S., Dietz, T., Duraiappah, A.K., Oteng-Yeboah, A., Pereira, H.M., Perrings, C., Reid, W.V., Sarukhan, J., Scholes, R.J., Whyte, A. (2009). Science for managing ecosystem services: beyond the Millennium Ecosystem Assessment. *Proceedings of the National Academy of Sciences USA* , 106(5), 1305–12.

Castellano, S., Balletto, E. (2002). Is the partial mantel test inadequate? *Evolution* , 56, 1871–1873.

Coissac, E., Riaz, T., Puillandre, N. (2012). Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular Ecology,*21(8), 1834–1847.

Coyne, J.A., Orr, H.A. (2004). Speciation. *Sinauer Associates* , Sunderland, MA

Craft, K.J., Pauls, S.U., Darrow, K., Miller, S.E., Hebert, P.D., Helgen, L.E., Novotny, V., Weiblen. G.D. (2010). Population genetics of ecological communities with DNA barcodes: An example from New Guinea Lepidoptera. *Proceedings of the National Academy of Sciences* , 107(11), 5041-5046.

Crispo, E., Bentzen, P., Reznick, D.N., Kinnison, M.T., Hendry, A.P. (2006). The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology,* 15, 49–62.

Dapporto, L., Cini, A., Vodă, R., et al. (2019). Integrating three comprehensive data sets shows that mitochondrial DNA variation is linked to species traits and paleogeographic events in European butterflies.*Molecular Ecology Resources* , 19, 1623–1636. https://doi.org/10.1111/1755-0998.13059

Dincă, V., Zakharov, E.V., Hebert, P.D.N., Vila, R. (2011). Complete DNA barcode reference library for a country's butterfly fauna reveals high performance for temperate Europe. *Proceedings of the Royal Society B* , 278 (1704), 347–355. http://doi.org/10.1098/rspb.2010.1089

Diniz-Filho, J., Bini, L., Hawkins, B. (2003). Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography* , 12, 53–64.

Diniz-Filho, J.A.F., Nabout, J.C., Telles, M.P.D.C., Soares, T.N., Rangel, T.F.L. (2009). A review of techniques for spatial modeling in geographical, conservation and landscape genetics. *Genetics and Molecular Biology,* 32, 203–211.

Diniz-Filho, J.A.F., Melo, D.B., de Oliveira, G., Collevatti, R.G., Soares, T.N., Nabout, J.C., de Souza Lima, J., Dobrovolski, R., Chaves, L.J., Veloso Naves, R., Dias Loyola, R., de Campos Telles, M.P. (2012). Planning for optimal conservation of geographical genetic variability within species. *Conservation Genetics* , 13(4), 1085–1093.

Diniz-Filho, J.A.F., Soares, T.N., Lima, J.S., Dobrovolski, R., Landeiro, V.L., de Campos Telles, M.P., Rangel, T.F., Bini, L.M. (2013). Mantel test in population genetics. *Genetics and Molecular Biology* , 36, 475–485.

Dirzo, R., Raven, P. (2003). Global state of biodiversity and loss.*Annual Review of Environment and Resources* , 28, 137–167.

Dobzhansky, T.G. (1937). *Genetics and the origin of species* . Columbia University Press, New York.

Domènech, M., Malumbres-Olarte, J., Enguídanos, A., Múrria, C., Arnedo, M.A. (2022). What DNA barcodes reveal: microhabitat preference, hunting strategy and dispersal ability drive genetic variation across Iberian spider species. *Insect Conservation and Diversity* , https://doi.org/10.1111/icad.12552

Dray, S., Legendre, P., Peres-Neto, P.R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling* , 196(3), 483-493.

Etzler, F.E., Wanner, K.W., Morales-Rodriguez, A., Ivie, M.A. (2014) DNA barcoding to improve the species-level management of wireworms (Coleoptera: Elateridae). *Journal of Economic Entomology* , 107, 1476–1485.

Frankham, R., Ballou, J.D., Briscoe, D.A. (2002). *Introduction to conservation genetics.* Cambridge: Cambridge University Press. doi:10.1017/CBO9780511808999

Freijeiro, A., Baselga, A. (2016). Spatial and environmental correlates of species richness and turnover patterns in European cryptocephaline and chrysomeline beetles. *ZooKeys* , 597, 81–99.

Fujisawa, T., Vogler, A. P., Barraclough, T. G. (2014). Ecology has contrasting effects on genetic variation within species versus rates of molecular evolution across species in water beetles. *Proceedings of the Royal Society B: Biological Sciences* , 282(1799), 20142476–20142476. doi:10.1098/rspb.2014.2476

Garcia-Raventós, A, Viza, A, Múrria, C. (2021). Taxonomic turnover and northward phylogenetic clustering reveal evidence for environmental filtering in structuring Trichoptera communities across Europe.*Freshwater Biology* , 66, 1060–1073. https://doi.org/10.1111/fwb.13699

García-Robledo, C., Erickson, D.L., Staines, C.L., Erwin, T.L., Kress, W.J. (2013). Tropical plant–herbivore networks: reconstructing species interactions using DNA barcodes. *PLoS One* , 8(1), e52967.

Gómez-Rodríguez, C., Freijeiro, A., Baselga, A. (2015). Dispersal and ecological traits explain differences in beta diversity patterns of European beetles. *Journal of Biogeography* , 42, 1526–1537. doi: 10.1111/jbi.12523

Goslee, S.C., Urban, D.L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* , 22(7), 1–19.

Gower, J.C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* , 53, 325–338.

Griffith, D.A., Peres-Neto, P.R. (2006). Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology* , 87, 2603–2613.

Guillot, G., Rousset, F. (2013). Dismantling the Mantel tests.*Methods in Ecology and Evolution,* 4, 336-344.

Hallmann, C.A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., et al. (2017). More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLoS ONE* , 12(10), e0185809. https://doi.org/10.1371/journal.pone.0185809

Harmon, L.J., Glor, R.E. (2010). Poor statistical performance of the Mantel test in phylogenetic comparative analyses. *Evolution* , 64, 2173–2178.

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R. (2003a). Biological identifications through DNA barcodes. *Proceedings. Biological Sciences* , 270, 313–321. doi:10.1098/rspb.2002.2218

Hebert, P.D.N., Ratnasingham, S., deWaard, J.R. (2003b). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences* , 270, S96–S99. doi:10.1098/rsbl.2003.0025

Hebert, P.D., Gregory, T.R. (2005). The promise of DNA barcoding for taxonomy. *Systematic Biology,* 54, 852–859.

Hedrick, P.W. (2001). Conservation genetics: where are we now?*Trends in Ecology and Evolution* , 16, 629–636.

Hendrich, L., Morinière, J., Haszprunar, G., Hebert, P.D., Hausmann, A., Köhler, F., Balke, M. (2015). A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3500 identified species to BOLD. *Molecular Ecology Resources* , 15, 795–818.

Hewitt, G.M. (2000). The genetic legacy of the Quaternary ice ages.*Nature* , 405, 907–913.

Hijmans, R.J. (2016). Geosphere: Spherical Trigonometry. *R package version* 1.5-5. https://CRAN.R-project.org/package=geosphere

Hof, C., Brändle, M., Brandl, R. (2006). Lentic odonates have larger and more northern ranges than lotic species. *Journal of Biogeography* , 33, 63–70. (doi:10.1111/j.1365-2699.2005.01358.x)

Hofreiter, M., Stewart, J. (2009). Ecological change, range fluctuations and population dynamics during the Pleistocene. *Current Biology* , 19, R584–R594.

Jackson, D.A. (1995). PROTEST: A PROcrustean randomization TEST of community environment concordance. *Ecoscience* , 2, 297–303.

Jackson, D.A. (2001). How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* , 129, 169–178.

Jelaska, L.Š. & Durbešić, P. (2009). Comparison of the body size and wing form of carabid species (Coleoptera: Carabidae) between isolated and continuous forest habitats. *Annales de la Société entomologique de France (N.S.),* 45:3, 327–338, DOI: 10.1080/00379271.2009.10697618

Joly, S., Davies, T. J., Archambault, A., Bruneau, A., Derry, A., Kembel, S. W., ... & Wheeler, T. A. (2014). Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. *Molecular Ecology Resources* , 14(2), 221–232.

Kenkel, N.C., Orloci, L. (1986). Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. *Ecology* , 67, 919–928. doi:10.2307/1939814

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* , 16, 111–120.

Koch, K. (1989). *Ökologie. - Die Käfer Mitteleuropas* . Band E1, 1–440. Krefeld: Goecke & Evers.

Koch, K. (1991). *Ökologie. - Die Käfer Mitteleuropas* . Band E2, 1–382. Krefeld: Goecke & Evers.

Koch, K. (1992). *Ökologie. - Die Käfer Mitteleuropas* . Band E3, 1–389. Krefeld: Goecke & Evers.

Köhler, F. & Klausnitzer, B. (Eds.) (1998). Verzeichnis der Käfer Deutschlands. *Entomologische Nachrichten und BerichteBeiheft* , 4, 1–185.

Köhler, F. (2000). Erster Nachtrag zum Verzeichnis der Käfer Deutschlands. *Entomologische Nachrichten und Berichte,* 44, 60–84.

Köhler, F. (2011a). 2. Nachtrag zum „Verzeichnis der Kafer Deutschlands“ (Kohler & Klausnitzer 1998) (Coleoptera). Teil 1. *Entomologische Nachrichten und Berichte,* 55, 109–174.

Kohler, F. (2011b). 2. Nachtrag zum „Verzeichnis der Kafer Deutschlands“ (Kohler & Klausnitzer 1998) (Coleoptera). Teil 2. *Entomologische Nachrichten und Berichte* , 55, 247–254.

Kohler, J., Rulik, B., Eberle, J., Kohler, F., Ahrens, D. (2022). Does ecological monitoring of saproxylic beetles benefit from the inclusion of larval stages? *Insect Conservation and Diversity*

Kuhne, G., Kosuch, J., Hochkirch, A., Schmitt, T. (2017). Extra-Mediterranean glacial refugia in a Mediterranean faunal element: the phylogeography of the chalk-hill blue *Polyommatus coridon*(Lepidoptera, *Lycaenidae*). *Scientific Reports* , 7, 43–533. http://doi.org/10.1038/srep43533

Kvist, S. (2013). Barcoding in the dark?: a critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge. *Molecular Phylogenetics and Evolution* , 69, 39–45.

Landeiro, V.L., Magnusson, W.E., Melo, A.S., Espirito-Santo, H., Bini, L.M. (2011). Spatial eigenfunction analysis in stream networks: do watercourse and overland distances produce different results? *Freshwater Biology* , 56, 1184–1192.

Lapointe, B.E. (1995). A comparison of nutrient-limited productivity in Sargassum natans from neritic vs. oceanic waters of the western North Atlantic Ocean. *Limnology and Oceanography* , 40, 625–633. doi:10.4319/lo.1995.40.0625

Lee, C.R., Mitchell-Olds, T. (2011). Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Molecular Ecology* , 20, 4631–464.

Legendre, P., Gallagher, E.D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia* , 129, 271–280.

Legendre, P., Legendre, L. (2012). Numerical Ecology, 3rd edition.*Elsevier* , Amsterdam, 990 pp.

Legendre, P., Bocard, D., Peres-Neto, P.R. (2005). Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monograph* , 75, 435–450.

Legendre, P., Fortin, M.J. (2010). Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources* , 10, 831–844.

Legendre, P., Fortin, M.J., Borcard, D. (2015). Should the Mantel test be used in spatial analysis? *Methods in Ecology and Evolution* , 6(11), 1239–1247.

Liebherr, J. (1988). Gene flow in ground beetles (Coleoptera: Carabidae) of differing habitat preference and flight-wing development.*Evolution* , 42, 129–137.

Lohse, G.A., Lucht, W.H. (1999). *Okologie. - Die Kafer Mitteleuropas,* Band 14: 1–395. Springer Spektrum Verlag, Berlin Heidelberg.

Magnacca, K.N., Brown, M.J. (2010). Mitochondrial heteroplasmy and DNA barcoding in Hawaiian Hylaeus (Nesoprosopis) bees (Hymenoptera:*Colletidae*). *BMC Evolutionary Biology* , 10, 174.

Manel, S., Schwartz, K., Luikart, G., Taberlet, P. (2003). Landscape genetics: Combining landscape ecology and population genetics.*Trends in Ecology and Evolution* , 18, 189–197.

Manel, S., Holderegger, R. (2013). Ten years of landscape genetics.*Trends in Ecology and Evolution* , 28, 614–621.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research* , 27, 209–220.

Moniz, M.B.J., Kaczmarska, I. (2010). Barcoding of diatoms: nuclear encoded ITS revisited. *Protist* , 161, 7–34.

Moore, W.S. (1995). Inferring phylogenies from mtDNA variation: Mitochondrial gene trees versus nuclear-gene trees. *Evolution* , 49, 718–726.

Murria, C., Iturrarte, G., Gutierrez-Canovas, C. (2020). A trait space at an overarching scale yields more conclusive macroecological patterns of functional diversity. *Global Ecology and Biogeography* , 29, 1729–1742. https://doi.org/10.1111/geb.13146

Nunn, C.L., Borgerhoff Mulder, M., Langley, S. (2006). Comparative methods for studying cultural trait evolution: a simulation study.*Cross-Cultural Research* , 40, 177–209.

Oden, N.L., Sokal, R. (1992). An investigation of three-matrix permutation tests. *Journal of Classification* , 9, 275–290.

Oksanen, J., Blanchet, F.G., Friendy, M., Kindt, R., Legendre, P., McGlin, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H. (2017). Vegan: community Ecology Package. *R package* version 2.4-2

Paliy, O., Shankar, V. (2016). Application of multivariate statistical techniques in microbial ecology. *Molecular Ecology* , 25, 1032–1057.

Papadopoulou, A., Bergsten, J., Fujisawa, T., Monaghan, M.T., Barraclough, T.G., Vogler, A.P. (2008). Speciation and DNA barcodes: testing the effects of dispersal on the formation of discrete sequence clusters. *Philosophical Transactions of the Royal Society B,* 363, 2987–2996.

Paradis, E., Claude, J., Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* , 20, 289–290.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* , 2(11), 559–572.

Pereira, H.M., Navarro, L.M., Martins, I.S. (2012). Global biodiversity change: the bad, the good, and the unknown. *Annual Review of Environment and Resources* , 37, 25–50. doi:10.1146/annurev-environ-042911-093511

Peres-Neto, P.R., Jackson, D.A. (2001). How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* , 129, 169–178.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ratnasingham, S., Hebert, P.D. (2007). BOLD: The Barcode of Life Data System (http://www. barcodinglife.org). *Molecular Ecology Notes* , 7, 355–364.

Ribera, I., Vogler, A.P. (2000). Habitat type as a determinant of species range sizes: the example of lotic–lentic differences in aquatic Coleoptera. *Biological Journal of Linnaean Society* , 71, 33–52. (doi:10.1111/j.1095-8312.2000.tb01240.x)

Ribera, I., Foster, G.N., Vogler, A.P. (2003). Does habitat use explain large scale species richness patterns of aquatic beetles in Europe?*Ecography* , 26, 145–152.

Rosindell, J., Hubbell, S.P., Etienne, R.S. (2011). The unified neutral theory of biodiversity and biogeography at age ten. *Trends in Ecology and Evolution* , 26, 340–348.

RStudio Team (2016). RStudio: Integrated Development for R.*RStudio* , Inc., Boston, MA URL

Rulik, B., Eberle, J., von der Mark, L., Thormann, J., Jung, M., Kohler, F., Apfel, W., Weigel, A., Kopetz, A., Kohler, J., Fritzlar, F., Hartmann, M., Hadulla, K., Schmidt, J., Horren, T., Krebs, D., Theves, F., Eulitz, U., Skale, A., Rohwedder, D., Kleeberg, A., Astrin, J.J., Geiger, M.F., Wagele, J.W., Grobe, P., Ahrens, D. (2017). Using taxonomic consistency with semi-automated data pre-processing for high quality DNA barcodes. *Methods in Ecology and Evolution* , 8, 1878–1887.

Rundell, R., Price, T.D. (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation.*Trends in Ecology & Evolution* , 24, 394–399.

Schmitt, T. (2009). Biogeographical and evolutionary importance of the European high mountain systems. *Frontiers in Zoology* , 6, 9. http://doi.org/10.1186/1742-9994-6-9

Schmitt, T. & Varga, Z. (2012). Extra-Mediterranean refugia: The rule and not the exception? *Frontiers in Zoology* , 9, 22. https://doi.org/10.1186/1742-9994-9-22

Schuldt, A., Assmann, T. (2011). Belowground carabid beetle diversity in the western Palaearctic – effects of history and climate on range-restricted taxa (Coleoptera, Carabidae). *ZooKeys* , 100, 461–474.

Šípek, P., Ahrens, D. (2011). Inferring larval taxonomy and morphology in Maladera species (Coleoptera: Scarabaeidae: Sericini) using DNA taxonomy tools. *Systematic Entomology* , 36, 628–643.

Sokal, R. (1979). Testing statistical significance of geographic variation patterns. *Systematic Zoology* , 28, 227–323.

Srivathsan, A., Meier, R. (2012). On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature.*Cladistics* , 28, 190–194.

Stewart, J.R., Lister, A.M., Barnes, I., Dalén, L. (2010). Refugia revisited: individualistic responses of species in space and time.*Proceedings of the Royal Society B: Biological Sciences* , 277, 661–671.

http://doi.org/10.1098/rspb.2009.1272

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* , 21, 2045–2050.

Toro, M., Caballero, A. (2005). Characterization and conservation of genetic diversity in subdivided populations. *Philosophical Transactions of the Royal Society B: Biological Sciences* , 360(1459), 1367–1378. http://doi.org/10.1098/rstb.2005.1680

Vellend, M., Lajoie, G., Bourret, A., Múrria, C., Kembel, S.W., Garant, D. (2014). Drawing ecological inferences from coincident patterns of population-and community-level biodiversity. *Molecular Ecology* , 23, 2890–2901.

Wagner, D.L., Grames, E.M., Forister, M.L., Berenbaum, M.R., Stopak, D. (2021). Insect decline in the Anthropocene: Death by a thousand cuts.*Proceedings of the National Academy of Sciences* , 118 (2), e2023989118; DOI: 10.1073/pnas.2023989118

Wang, C., Szpiech, Z.A., Degnan, J.H., Jakobsson, M., Pemberton, T.J., Hardy, J.A., Singleton, A.B., Rosenberg, N.A. (2010). Comparing spatial maps of human population-genetic variation using Procrustes analysis.*Statistical Applications in Genetics and Molecular Biology* , 9, 13.

Wang, I.J., Summers, K. (2010). Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Molecular Ecology* , 19, 447–458.

Wang, I.J., Glor, R.E., Losos, J.B. (2013). Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters* , 16, 175–182.

Wright, S. (1943). Isolation by distance. *Genetics* , 28, 114–138.

Yu, D.W., Ji, Y., Emerson, B.C., Wang, X., Ye, C., Yang, C., Ding, Z. (2012). Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* , 3, 613–623.

Table 1. Number and their respective percentages (%) of species resulting to be significant in the correlation analysis of infraspecific distances vs. geographic distances, as performed by the Mantel test, NMDS+PCA, and PCNM+PCA (in comparison to the total species number of each category; N_{spec}), respective selected life traits of the species (feedings style, habitat, body size and biotope).

	$\mathbf{N_{spec}}$	**Mantel**	**%**	**NMDS+ PCA**	**%**	**PCNM+ PCA**	**%**
Feeding style							
Coprophagous	29	8	27.59	6	20.69	5	17.24
Polyphagous	15	4	26.67	3	20.00	1	6.67
Mycetophagous	156	27.3	17.50	14	8.97	13	8.33
Necrophagous	27	3	11.11	0	0	1	3.70
Phytophgous	585	80	13.68	60	10.26	40	6.84
Saprophagous	99	17	17.17	9	9.09	8	8.08
Xylophagous	200	32.3	16.15	19	9.50	14	7.00
Zoophagous	674	101	14.99	55	8.16	39	5.79
Habitat							
Soil	398	60	15.08	36	9.05	26	6.53
Eurytopic	22	3	13.64	0	0	0	0
Rotten matters	167	21	12.57	12	7.19	9	5.39

	$\mathbf{N_{spec}}$	**Mantel**	**%**	**NMDS+ PCA**	**%**	**PCNM+ PCA**	**%**
Nest	28	2	7.14	4	14.29	3	10.71
Mushrooms	29	5	17.24	1	3.45	2	6.90
Dead woods	404	41	10.15	39	9.65	32	7.92
Vegetation	632	88	13.92	68	10.76	41	6.49
Wetlands	105	30	28.57	7	6.67	8	7.62
Biotope							
Eurytope	305	41	13.44	22	7.21	17	5.57
Wetland	369	75	20.33	30	8.13	26	7.05
Open land	420	56	13.33	52	12.38	30	7.14
Forest	691	78	11.29	63	9.12	48	6.95
Body size							
xs	594	77	12.96	47	7.91	33	5.56
s	626	99	15.81	63	10.06	49	7.83
m	386	53	13.73	40	10.36	22	5.70
l	157	18	11.46	15	9.55	13	8.28
xl	22	3	13.64	3	13.64	4	18.18
Overall mean			**15.57**		**9.44**		**7.50**

Table 2. Comparison between the three analytic methods. Only significant scores resulting from the multi-linear models are reported. The first row of each eco-class shows the F-statistic score (when significant) for the correspondent pair of variables. A dash indicates no support by the models (i- intercept).

	Mantel	**Mantel**	**Mantel**	**NMDS+PCA**	**NMDS+PCA**	**NMDS+PCA**	**P**
ANOVA	**Statistic**	**Signifi-cance**	**Binary Sig.***	**Statistic**	**Signifi-cance**	**Binary Sig.**	**S**
Feeding style	-	3.5	-	-	-	-	-
coprophagous (i)	0.21	0.4	-6.51	0.67	0.57	-10.39	0
polyphagous	-	-	-	-	-	-	-
mycetophagous	-	-	-	-	-	-	-
necrophagous	-	-	-	-	-	-	-
phytophagous	-	-	-	-	-	-	-
saprophagous	-	-	-	-	-	-	-
xylophagous	-	-	-1.33	-	-	-	-
zoophagous	-	-	-	-	-	-	-
Habitat	6.9	4.9	-	2.8	2.9	-	-
soil (i)	0.27	0.41	-7.57	0.64	0.52	-11.17	0
eurytopic	-	-	-	-	-	-	-
rotten matters	-	-	-	-	-	-	-
nest	-	-	-	-	-	-	-
mushrooms	-	-	-	0.08	-0.11	-	-
dead woods	-0.06	0.05	-	-	-	-	-
vegetation	-0.06	-	-	0.03	-0.05	-	0
wetlands	0.11	-0.07	-	-	-	-	-
Biotope	19.9	17.1	-	-	3	-	-
eurytopic (i)	0.2	0.46	-6.93	0.64	0.52	-11.15	0
wetland	0.1	-0.05	-	-	-	-	-

	Mantel	**Mantel**	**Mantel**	**NMDS+PCA**	**NMDS+PCA**	**NMDS+PCA**	**P**
open land	-	-	-	-	-	-	-
forest	-	-	-0.62	-	-	-	-
Body size	-	-	-	-	-	-	-
l (i)	0.22	0.44	-7.5	0.65	0.5	-10.98	0
m	-	-	-	-	-	-	-
s	-	-	-	-	-	-	-
xl	-	-	-	-	-	-	-
xs	-	-	-	-	-	-	-
Distance classes	8	4.2	-	3.5	4.9	-	1
>300 (i)	0.21	0.54	-7.85	0.62	0.53	-10.69	0
100	-	-	-	-	-	-	-
150	-	-	-	-	-	-	-
200	-	-	-	-	-	-	-
300	-	-	-	-	-	-	-

*NA is reported in correspondence of the GLM predictor (Binary sig.) as no analysis of variance had been computed for such categorical variable.

Figure 1. Map of the sampling sites of specimens considered for this study across the Central Europe.

Figure 2. Flow chart illustrating the two alternative analyses to the Mantel test used to detect the intraspecific relationship between geographical distances and genetic distances. Left side: PCA method; Middle: the NMDS method; it starts with the same matrix and ends with a set of axes representing the DNA distances. Right side: PCNM technique performed on geographic coordinates. The PCNM produced a set of eigenvectors which had been regressed in Procrustes superimposition analysis first against NMDS axis, then against PCA eigenvectors. Finally, significance test with Protest.

Figure 3. Mean intraspecific genetic distances plotted vs A) mean geographic intraspecific distances (in km); B) number of sampled individuals per species (N_ind); C) number of sampling sites per species (loc); D) vs geographic distances (km) for guilds with different feeding styles; E) vs geographic distances (km) for guilds with different habitat preferences; F) vs geographic distances (km) for body size classes; G) vs geographic distances (km) for guilds of biotope preference.

Figure 4 . Violin plots and bar plots of statistic scores from Mantel, NMDS+PCA, and PCNM+PCA analyses across the different ecological guilds and subcategories, differentiated for species with significant (dark grey) and non- significant (light grey) correlation.

Habitat preferences: h_b =soil, h_e=eurytop, h_f= rotting matters, h_n=nest, h_p=vegetation, h_t=dead wood, h_v=vegetation; h_w= water; biotope preferences: b_e= eurytopic; b_f =wetlands ; b_o= open-land biotypes; b_w=forests; body size classes: s_xs= extra small; s_s= small; s_m=medium; s_l=large and s_xl= extra-large; feeding style: f_c = coprophagous; f_e = polyphagous; f_m= mycetophagous; f_n= necrophagous;f_p= phytophagous; f_s= saprophagous; f_x= xylophagous; f_z= zoophagous; geographical distance classes: highest treashold measured in km.

Supplement Figure 1. Distribution of the number of specimens (above) and species (middle) on the five classes of geographical distances and the amount of sampling localities (below).

Supplement Figure 2. Bar plots (above) showing the number of species across the ecological guilds. Violin plots (below) illustrate the number of specimens per species for each considered ecological trait: habitat preference (A); feeding style (B); biotope preference (C); body size classes (D); the width of violin indicates the total number of such cases.

Supplement Figure 3. Relationships between pairwise geographic and genetic distances across guilds for each ecological variable.

Supplement Figure 4. Dot plot showing the relationship between the sampling variables: number of localities (loc); number of individuals per species (N_ind); mean intraspecific geographic distance (km).

Supplement Figure 5. Significance levels for Mantel, NMDS, and PCA tests (y-axis: significance value; x-axis: p-value). On the left panel, significant species are shown in detail; right panel: all the examined species that showed a positive and significant relationship for at least one test.

Supplement Figure 6. Comparison between the three correlation methods. on the bottom-left are regressed the three technique Statistic scores. On the top-left are regressed the significance scores.

Supplement Figure 7. Histogram showing the 1785 species distribution across the 95 examined families. In the chart all the species resulting significant in at least one of the three statistical tests performed during the study have been filled in dark grey.

Supplement File S1

Tab. S1. Metadata list including all the examined species and their relative ecological classes appartenence.