Title

Factors affecting the efficiency of molecular species delimitation in a species-rich insect family

Running Title

Molecular species delimitation efficiency

Authors

Giulia Magoga[*1], Diego Fontaneto[2], Matteo Montagna[*1,3]

Authors affiliations

[1]Dipartimento di Scienze Agrarie e Ambientali - Università degli Studi di Milano, Via Celoria 2, 20133, Milano, Italy.

[2]Consiglio Nazionale delle Ricerche (CNR), Istituto di Ricerca Sulle Acque (IRSA), Molecular Ecology Group (MEG), Largo Tonolli 50, 28922, Verbania, Italy.

[3]BAT Center - Interuniversity Center for Studies on Bioinspired Agro-Environmental Technology, University of Napoli "Federico II", Via Università 100, 80055, Portici, Italy.

**corresponding authors:* giulia.magoga@unimi.it, matteo.montagna@unimi.it

Abstract

In the contest of global biodiversity loss, molecular species delimitation approaches can be very useful for accelerating species discovery through DNA taxonomy and inventory through DNA metabarcoding. In this study, the effect of some intrinsic factors on the efficiency of various single-marker species delimitation methods (fixed and variable nucleotide distance thresholds,

ABGD, ASAP, GMYC, mPTP) was tested on more than 90 empirical datasets, derived from a set of 7,237 COI sequences attributed to 542 leaf beetles species (Coleoptera: Chrysomelidae). The considered factors were: i) the number of haplotypes per species (as a proxy for genetic diversity); ii) the geographic distance among conspecific collection localities (as a proxy of sampling width); iii) the difficulty related to morphological identification of species; iv) the taxonomic rank. Distance-based methods, with on average more than 70% of match with morphological identification, outperformed those relying on phylogenetic trees, with less than 59%. A high number of haplotypes per species was found to have a negative effect on delimitation efficiency, whereas large geographic distances within species had a positive effect. All methods delimitations (except for GMYC) were significantly affected by the presence of species that are difficult to be identified, decreasing their efficiency. Finally, the only method influenced by the taxonomic rank of the dataset was GMYC, showing lower efficiency in datasets at the genus than at higher levels. The observed biases we highlighted affecting efficiency could be accounted for when developing input datasets for species delimitation analyses to obtain a more reliable representation of biological diversity.

Keywords

Chrysomelidae; ABGD; ASAP; GMYC; mPTP; Cytochrome c oxidase subunit I COI;

Introduction

In the context of the rapid impoverishment of global biodiversity (Hallmann et al., 2017; Sánchez-Bayo & Wyckhuys, 2019; Van Klink et al., 2020), it is easy to understand the reason of the success of molecular taxonomy that, exploiting molecular information, has the potential to accelerate the identification of organisms and the discovery process of new taxa (Hebert &

Gregory, 2005; Swartz, Mwale, & Hanner, 2008; Monaghan et al., 2009; Mutanen, Kaila, & Tabell, 2013; Gaytán et al., 2020). The DNA-based identification methods represented by DNA taxonomy, applicable also in DNA metabarcoding (Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012), stand out as widely used approaches for biodiversity surveys (e.g., Telfer et al., 2015; Pavan-Kumar, Gireesh-Babu, & Lakra, 2015; deWaard et al., 2019). These methods, using standardized gene regions, allow the speeding up of organism identification and help overcoming issues in morphological taxonomy, for example by making it possible to detect the presence of organisms merely from their DNA, released in the environment (e.g., Ficetola, Miaud, Pompanon, & Taberlet, 2008; Montagna et al., 2018; Ruppert, Kline, & Rahman, 2019). Beside molecular identification of already known species, species delimitation tools in DNA taxonomy allow inferring hypothetic species and/or evolutionary significant units from molecular data (Puillandre, Lambert, Brouillet, & Achaz, 2012; Tang, Humphreys, Fontaneto, & Barraclough, 2014). Species delimitation molecular methods are used for both biodiversity investigation without a priori hypothesis on the possible species (Dincă et al., 2015; Gómez-Zurita, 2016) or, more frequently, as support to resolve taxonomic issue when other delimitation approaches give uncertain results (e.g., Montagna et al., 2016; Kajtoch, Montagna, & Wanat, 2018; Plewa et al., 2018; García-Melo et al., 2019), in the framework of the so-called "integrative taxonomy" approach (Dayrat, 2005). Several molecular species delimitation methods have been proposed: early methods analysed alloenzymes variability, allele frequency or nuclear genes co-dominance (Highton, 1989; Porter, 1990; Davis & Nixon, 1992; Good & Wake, 1992; Highton, 2000), but, currently, the most widely adopted molecular delimitation methods rely on DNA sequences as markers. The result of species delimitation analyses using nucleotide sequences is more accurate when taking information from more than one marker (Rubinoff & Holland, 2005; Dupuis, Roe, & Sperling, 2012); nevertheless, single-locus species

delimitation approach is widely applied on data derived from DNA barcoding studies (DeSalle & Goldstein, 2019; Puillandre, Brouillet, & Achaz, 2020) and is still the only reliable approach for metabarcoding. Even when multiple markers are used for metabarcoding, the origin of the different markers cannot be considered from the same individual and any delimitation analysis has to be performed independently for each marker. Molecular taxonomy approaches are particularly useful for investigating diversity of those organisms for which morphological identification is challenging such as bacteria, fungi but also invertebrates (Göker, García-Blázquez, Voglmayr, Tellería, & Martín, 2009; Seifert, 2009; Ceccarelli, Sharkey, & Zaldívar-Riverón, 2011). For most invertebrates, the standard marker for DNA barcoding is a fragment at the 5' region of Cytochrome c oxidase I (COI), which is considered robust and reliable due to a nucleotide variability that allows the discrimination of organisms at species level and, in particularly, due to the highly consolidate primers for its amplification, suitable for the majority of the groups (Hebert, Cywinska, Ball, & De Waard, 2003; Mioduchowska, Czyż, Goødyn, Kur, & Sell, 2018). One of the first methods proposed for delimiting species starting from COI barcodes is a phenetic approach that relies on sequences pairwise nucleotide distances (Sneath & Sokal, 1973). The approach is based on the idea that a "barcoding gap" in the frequency distribution of intra- and inter-specific distances of different taxa occurs at the same value of distance (Hebert et al., 2003), allowing to define a fixed threshold to discriminate between intra- and inter-specific nucleotide variability (Hebert et al., 2003). This approach has been largely criticized since a fixed threshold value may not be suitable for delimiting species when taxa with different evolutionary histories are analysed together (e.g., Meyer & Paulay, 2005). In light of this, the application of group-specific *ad hoc* thresholds is supposed to result in more accurate delimitations. Without any a priori species hypothesis, the *ad hoc* threshold for delimiting a group of taxa can be estimated looking for the barcoding gap in the frequency

distribution of nucleotide distances, but since a real gap infrequently occurs, a minimum in the distribution can be identified as a point of transition between intra- and inter-specific level. According to this idea, the Automatic Barcode Gap Discovery method (ABGD; Puillandre et al., 2012) was developed. ABGD stand out as one of the most used species delimitation methods (more than 1,800 citation in Google Scholar up to January 2021) and, very recently, from the same authors a new method named ASAP was released (Puillandre et al., 2020). ASAP was developed to be even more user-friendly than ABGD, improving the choice of priors and of the ultimate delimitation, which in ABGD are left to the user. Alternative methods to distance-based one, explicitly accounting for evolutionary processes (Hickerson, Meyer, & Moritz, 2006), are represented by the phylogenetic and coalescent-based methods; among them, the most used ones are the Generalized Mixed Yule Coalescent model (GMYC, Pons et al., 2006; Fujisawa & Barraclough, 2013) with >1700 citations in Google Scholar up to January 2021, its Bayesian implementation bGMYC (Reid & Carstens, 2012), the Poisson Tree Process (PTP, Zhang, Kapli, Pavlidis, & Stamatakis, 2013) with >1200 citations, and its multi-rate extension (mPTP, Kapli et al., 2017), in addition to other approaches that have not been so frequently used (e.g. K/theta by Birky, Adams, Gemmel, & Perry, 2010; haploweb by Flot, Couloux, & Tillier, 2010; both with <120 citations). GMYC and PTP require as input a phylogeny of taxa estimated from DNA sequences: GMYC uses the topology of a tree to identify the maximum likelihood solution separating the branches between species modelled by a Yule process and the branches within species modelled by neutral coalescent (Fujisawa & Barraclough, 2013); PTP finds the transition point between intra- and inter-specific processes assuming a two parameter model that accounts for speciation and for the coalescent process based on the Poisson distribution of branch lengths (Zhang et al., 2013; Kapli et al., 2017).

Despite the fact that DNA-based species delimitation methods are widely used and

consolidated, factors intrinsic to the analysed data that could lead to improper delimitation results still need to be empirically investigated. Some studies were carried out for this purpose, but the majority of them were performed on simulated data where dataset characteristics were a priori defined in order to test for specific hypotheses. Among these tests, some evaluated the effect of sampling scale, variation of the effective population size, of speciation rate and mutation rate, marker length and sampling size, on GMYC, ABGD and PTP performances (Lohse, 2009; Reid & Carstens, 2012; Esselstyn, Evans, Sedlock, Anwarali Khan, & Heaney, 2012, Fujisawa & Barraclough, 2013; Dellicour & Flot, 2018). However, when analysing real data, almost none of the previous factors can be realistically accounted for. Testing the factors affecting species delimitation on real data can help to define practical guidelines to establish the most reliable experimental design and/or to account for biases when reading delimitation results. Some studies in this direction are available (Talavera, Dinca, & Vila, 2013, Ahrens et al., 2016, Pentinsaari, Vos, & Mutanen, 2017), but further factors and methods should be considered to obtain a complete overview on the potential biases of the methods.

The aim of the study is to analyse the influence of data-related factors on the efficiency of six molecular species delimitation methods (i.e., fixed nucleotide distance threshold, *ad hoc* developed nucleotide distance threshold, ABGD, ASAP, GMYC, and mPTP). A collection of ~7,200 COI sequences belonging to 542 Euro-Mediterranean leaf beetles species (Coleoptera: Chrysomelidae) was investigated. The sequences are representative of almost ¼ of Euro-Mediterranean Chrysomelidae fauna, a family including almost entirely phytophagous species, but with highly different ecological traits regarding habitat (e.g., Brunetti, Magoga, Iannella, Biondi, & Montagna, 2019, Mende, Biström, Meichssner, & Kölsch, 2010), trophic specialization (Mardulyn, Milinkovitch, & Pasteels, 1997; Fuss, Geiser, & Patzner, 2005), range of distribution (Schmitt & Rönn, 2011; Biondi, Urbani, & D'Alessandro, 2013), size and

dispersal abilities (Strauss, 1988; Mende et al., 2010; Piper & Compton, 2010). Chrysomelidae COI sequences were assembled in datasets with different characteristics to test the impact on molecular delimitation efficiency of the following data-related factors: i) the number of haplotypes per species, as a proxy for species *genetic diversity* (and potentially also for sample size). A higher number of haplotypes per species could have differential impacts on species delimitation methods: on the one hand genetic variability will be better represented (Goodall-Copestake, Tarling, & Murphy, 2012), but on the other hand it could blur the differences between species. ii) The *sampling structure* evaluated considering geographic distances between collection localities of conspecifics. Sampling structure could impact species delimitation efficiency, especially when genetic and geographic distances are related (Mason, Fletcher, Gill, Funk, & Zamudio, 2020); a discontinuous sampling could cause oversplit (Lohse, 2009; Talavera et al., 2013; Mason et al., 2020), a limited sampling could favour delimitation efficiency (Pentinsaari et al., 2017), whereas wider sampling could result in both split of species or better delimitation possibly depending on species population structure. iii) The *taxonomic rank* of the dataset, in this work considered at the genus, subfamily and family level, could be viewed as a proxy of the overall phylogenetic distance among the analysed species. Having only closely related sequences or also sequences from rather distant species in a single dataset may affect the estimation of several parameters in the methods from DNA taxonomy, including also reliability of the phylogenetic reconstructions. iv) The *morphological distinctiveness* of the species, evaluated as the difficulty related to their morphological identification. Morphological identification could be easy in Chrysomelidae, with clear and unambiguous features for some groups of species but in other cases only subtle differences of the genital apparatuses are present as diagnostic characters (e.g., Bezdek & Baselga, 2015; Montagna et al., 2016; Magoga, Coral Sahin, Fontaneto, & Montagna, 2018). Closely related

169 species, with only subtle differences in the copulatory system, may have recently diverged, with
170 no time for a clear differentiation between species in the molecular marker chosen for molecular
171 identification.
172 We explore the variability in the efficiency of the different methods for molecular identification
173 of species comparing the match with morphological identification, here considered as the most
174 appropriate descriptor of diversity in this well-studied group of insects (Doguet, 1994;
175 Warchałowski, 2003; Nadein, 2013; Petitpierre, 2016; Nie et al., 2019; Petitpierre, 2019),
176 disentangling the effects of the potential confounding factors we described including genetic
177 diversity, sampling structure, taxonomic rank, and morphological distinctiveness.
178
179 ***Material and Methods***
180 *Datasets*
181 In the present work the collection of COI sequences of Chrysomelidae (Insecta: Coleoptera)
182 developed and analysed in Magoga et al. (2018) was used as reference for the performed
183 analyses. This COI collection consists of 7,237 sequences (average 652 bp; range: 460 bp to
184 658 bp) that are taxonomically assigned to 542 species. The Magoga et al. (2018) COI sequence
185 collection was divided into multiple datasets according to taxa membership to the taxonomic
186 level of family, subfamily and genus (Fig. 1A); then the datasets were aligned at codon level
187 using MUSCLE in MEGA X (Kumar, Stecher, Li, Knyaz, & Tamura, 2018). The obtained
188 alignments were used as input for the species delimitation analyses, *per se*, after outgroups
189 inclusion, or after the inference of a phylogenetic tree (see below). The datasets used to infer
190 phylogenetic trees, input of tree-based species delimitation methods, were collapsed with R
191 software 3.6.2 (R Core Team, 2019) in order to retain for each species only unique haplotypes.
192

Molecular species delimitation analyses

In the present study, the following nucleotide distance-based and tree-based molecular species delimitation methods were adopted (Fig. 1A): *i*. the 3% nucleotide distance threshold proposed by Herbert et al. (2003); *ii*. nucleotide distance thresholds *ad hoc* estimated on each dataset; *iii*. ABGD (Puillandre et al., 2012); *iv*. ASAP (Puillandre et al., 2020) *v*. GMYC (Pons et al., 2006; Fujisawa & Barraclough, 2013); *vi*. the Multi-rate Poisson tree processes (mPTP; Kapli et al., 2017). The 3% nucleotide distance threshold and *ad hoc* threshold analyses were carried out on 96 datasets (one family, 12 subfamily, and 83 genus datasets). ABGD and ASAP analyses were performed on 94 datasets (one family, 12 subfamily, and 81 genus datasets) since datasets composed by less than three sequences were excluded from the analyses. GMYC and mPTP analyses were performed on phylogenetic trees inferred on 92 datasets (one family, 12 subfamily, and 79 genus datasets) by excluding two additional genus datasets with only one sequence per species after haplotype collapsing.

The K2P pairwise nucleotide distance matrices, required for the species delimitation through 3% nucleotide distance threshold and *ad hoc* nucleotide distance threshold methods, were estimated for each dataset using the R software library APE 5.3 (Paradis, Claude, & Strimmer, 2004). The *ad hoc* nucleotide distance threshold for each dataset was estimated using R function *localMinima* of package SPIDER 1.5.0.9000 (Brown et al., 2012). The function, based on the barcoding gap concept, identifies the minima in the density of nucleotide distances as possible thresholds for delimiting the species present in the analysed dataset. The R function *tclust* of package SPIDER was subsequently used to cluster nucleotide sequences at the distance threshold of 3% and at the *ad hoc* threshold value previously identified.

ABGD analyses were performed using the command-line version downloaded from https://bioinfo.mnhn.fr/abi/public/abgd/ with the following settings: K2P nucleotide

substitution model (Kimura, 1980) to infer the nucleotide distance; relative gap width of 1.5, when gap was not found using this value, a width of 1 and 0.5 were set; prior P ranging from 0.001 to 0.1 and the remaining parameters were left as default. ASAP analyses were run using the program web-interface (https://bioinfo.mnhn.fr/abi/public/asap); K2P (Kimura, 1980) was selected as nucleotide substitution model and other parameters were left as default. ASAP delimitation was defined evaluating both the partitions with first and the second best asap-score according to Puillandre et al., 2020.

In order to perform GMYC and mPTP species delimitation analyses, an ultrametric tree was inferred using the software BEAST v 1.8 (Drummond, Suchard, Xie, & Rambaut, 2012) from each haplotype-reduced dataset (Tang et al., 2014). One or more (in the case of the family dataset) orthologous sequences from the appropriate outgroup was included to properly root the tree. For each dataset two to five independent MCMC runs were performed using the following parameters: Markov chain length from $30*10^6$ to $300*10^6$ generations depending on the runs convergence assessed examining the estimated sample size of each parameter of the model and a visual inspection of the likelihood with TRACER (Drummond et al., 2012); sampling of trees and parameters every 1,000 to 5,000 generations depending on the total number of generations; models of nucleotide evolution as selected according to the Bayesian information criterion after the analysis performed by jModelTest 2 (Darriba, Taboada, Doallo, & Posada, 2012); Yule process as speciation model (Yule, 1925); all other priors set to their default values. The runs were then pooled or resampled according to the number of performed generations, after removal of the proper tree burn-in fraction, using LogCombiner (Drummond et al., 2012) and the majority-rule consensus tree obtained by TreeAnnotator (Drummond et al., 2012).

Single-threshold GMYC species delimitation analyses were performed using the default settings of the R function *gmyc* of the library SPLITS 1.0.19 (Ezard, Fujisawa, & Barraclough,

2009).

mPTP analyses were performed through the binary version 0.2.4 available on https://github.com/Pas-Kapli/mptp. For each dataset, we performed ten different runs with the following settings: mcmc run of $100*10^6$ generations (steps), sample frequency every 5,000 generations and a burnin of 20,000 generations; the convergence of the independent runs was assessed through the average standard deviation of delimitation support values (ASDDSV) and the overall support for the ML estimate calculated computing the mean of the average support values (ASV) over the ten runs.

The delimitation results (called units) obtained for each analysis were compared to the identity of the morphological species and classified in the following categories adopting an *ad-hoc* developed R script: i) *match* = all the sequences of the same morphological species were delimited as belonging to the same unit; ii) *split* = the sequences of a species are delimited as belonging to two or more units; iii) *merge* = the sequences of two or more species are included in the same unit; and iv) *mixture* = some sequences of a species are split while others are merged (Fig. 1A).

Factors affecting species delimitation

In this work the number of observed matches in each analysis was used as a proxy of the achieved efficiency. Efficiency of the different species delimitation methods was defined as the number of matches (successes) against the number of failures, counted as the sum of merges, splits, and mixtures.

The factors potentially affecting species delimitation were represented by a series of data-related variables (Fig. 1B). The factors took in account: *i*. the mean number of haplotypes per species in a dataset (genetic diversity and sample size); *ii*. the median value of the geographic

distance between conspecific collection localities (sampling structure and width); *iii*. the difficulty level of the morphological identification of the species (morphological distinctiveness of species); *iv*. the taxonomic rank of the dataset.

The number of haplotypes was calculated for each species using R and then for each dataset the average number estimated. The geographic distances between collection localities were calculated with the R package GEOSPHERE 1.5.10 (Hijmans, 2017) starting from the geographic coordinates of the sampling points in WGS84 system (Table S1). The median value for each species was then extracted and for each dataset, the median of these values was calculated. As already mentioned, the analysed datasets belong to different taxonomic ranks (i.e.,family, subfamily, and genus); these ranks were accounted to evaluate their influence on delimitation efficiency. In order to assess whatever there is an effect of morphological distinctiveness, each species present in this study was assigned to one of the following three categories of morphological identification difficulty: *i*. species that are easy to be morphologically identified due to the presence of clear and easily detectable diagnostic characters (level I); *ii*. species that can be identified through subtle differences in morphological characters (excluding genitalia) or on the basis of strong differences in the genitalia shape (spermatheca or the median lobe of the aedeagus) (level II); and *iii*. species that are identifiable only on the basis of subtle differences in the shape of genitalia (level III). The level of difficulty was assigned to species using information from specialized literature reporting species descriptions and dichotomous keys (e.g., Müller, 1953; Burlini, 1955; Doguet, 1994; Warchałowski, 2003, Petitpierre, 2016; Petitpierre, 2019). For each dataset, the proportion of species assigned to each difficulty level was calculated and registered in two variables, namely proportion of difficult species (as level II + level III) and proportion of extremely difficult species (only level III), in order to assess the effect of the increasing difficulty.

289

290 *Statistical analyses*

291 Before performing any explicit test, the presence of correlation between the explanatory
292 variables was assessed through the R library PSYCH 1.9.12.31 (Revelle, 2018).

293 A generalized linear mixed-effects model (GLMEM) was then run for testing the differential
294 influence of the explanatory variables that passed the preliminary test of multiple correlation
295 (namely number of haplotypes, geographic distance, taxonomic level, taxonomic difficulty and
296 taxonomic extreme difficulty) and the effect of species delimitation method on efficiency. The
297 model was fitted using the R package LMERTEST 3.1.2 (Kuznetsova, Brockhoff, & Christensen,
298 2017) using a negative binomial distribution given that the response variable was proportional
299 data, and adding the identity of dataset as a random effect in the error structure of the model to
300 remove the effect of pseudoreplication, given that the same datasets were analysed with
301 different methods. In addition, the model included the interaction terms of each explanatory
302 variable with the species delimitation method.

303 In order to understand the differences between methods, additional statistical models were run
304 separately for each delimitation method using GLMs with the same structure of explanatory
305 variables of the previous GLMEM (obviously, excluding method).

306 The datasets of individuals at the genus, subfamily and family level are nested, but the data
307 used in the models are not nested, because the analyses were performed separately for each
308 taxonomic rank; thus, the nested effect of the taxonomic ranks was not included as a random
309 effect in the error structure of the models.

310 The outcomes of GLMEM and GLMs are reported in the results section as Type II analysis-of-
311 variance tables fitted with the R package CAR 3.0.6 (Fox & Weisberg, 2019).

312

Results

Datasets composition and species delimitation efficiency

The set of 7,237 sequences was organized into 99 datasets according to the taxonomic levels, resulting in one family, 12 subfamily, and 86 genus datasets. Three genus datasets were excluded from the delimitation analyses since they were composed by only one COI sequence. The datasets were highly different in term of species and COI sequences number ranging from the family dataset composed by 542 species and 7,237 sequences (4,066 after haplotype reduction) to a mean of 6.5 species [range 1:93] and 87 sequences [2:1,014] for the genus datasets (50.1 sequences after haplotypes reduction in mean) (Table S2).

Overall, 64.4% of the species were delimited in the same category by all the four delimitation methods when the family dataset is considered, 63.8% in the case of subfamily datasets, and only 50.9% at the genus level.

The highest percentage of matches between molecular delimited units and morphospecies (80.5% on average) was obtained delimiting species through the *ad hoc* nucleotide distance thresholds (Table S3, Table S4). Molecular delimitation using ABGD and ASAP resulted in a similar percentage of matches (77.6% and 77.9% on average, respectively), moderately higher than using the 3% nucleotide distance threshold (72.8% on average) (Fig. 2A). Ad hoc nucleotide distance threshold, ABGD and ASAP showed also a nearly equal efficiency at any taxonomic level (Fig. 2B). The likelihood-ratio test of GMYC analyses rejected the null model (p-value < 0.05) for 51 out of 94 datasets, thus being unable to delimit molecular taxonomic units in almost half of the datasets (43 datasets, Table S4); the average percentage of matches resulted of 33.2% for this method (Fig. 2A). In mPTP analyses, ASDDSV values resulted < 0.01 for 90 over 94 datasets, indicating the convergence of the ten independent MCMC runs. For the majority of the datasets (83 over 94) ASV values resulted high (median: 92.4%),

337 suggesting that the ML solution was supported by the data. For the 11 datasets with ASV values
338 lower then 50%, the results were considered not significant (Table S4). On average this method
339 produced 58.7% of match (Fig. 2A).
340 The highest number of merged morphological species was observed for GMYC, while split
341 species were more common from mPTP and 3% distance threshold delimitation (Fig. 2A). The
342 highest number of mismatches was associated to the genus datasets delimited by tree-based
343 methods, GMYC and mPTP (Fig. 2B). Delimitation results obtained on each dataset with each
344 method are reported in Table S4.
345
346 *Explanatory variables*
347 *Number of haplotypes.* Within the analysed datasets on average ~6 haplotypes per species were
348 present (range 1:40.5; Table S2).
349 *Geographic distance between conspecific collection localities*. A median distance among
350 collection localities of conspecifics of ~219 km was obtained with values ranging from 0
351 (species collected in one locality only) to 1,685 km (Table S2). The highest value was observed
352 for *Tituboea biguttata* for which two specimens were collected one in Italy and one in Morocco.
353 *Difficulty in species morphological identification.* About 35% of the species included in the set
354 of COI sequences analysed were easily identifiable morphologically (level I), 53.5% were
355 considered of intermediate difficulty (level II), 11.5% were extremely difficult (level III).
356 Among the species categorized in level I there are for example those of the genera *Zeugophora*
357 and *Syneta* and of the subfamily Hispinae; all the species included in level II belong to 11
358 genera (e.g., *Neocrepidodera*, *Orestia*, *Plagiosterna*, *Plateumaris*); most of the species of
359 *Altica* and *Oulema* were assigned to level III (Table S2).
360

Factors affecting efficiency in species delimitation

The five explanatory variables (mean number of haplotypes, median geographical distance, proportion of difficult species, proportion of extremely difficult species, taxonomic rank) were not correlated among each other (Fig. S1) and were kept in the analyses.

The efficiency of species delimitation was significantly different between the six molecular species delimitation methods (GLMEM: $\chi^2 = 74.71$, $p < 0.0001$), with a complex scenario of different factors differentially affecting the efficiency of the methods (Table 1).

Analysing each method separately, all factors were found to significantly affect species delimitation efficiency (Table 2). The number of haplotypes per dataset resulted to affect the efficiency of 3% nucleotide distance threshold, *ad hoc* nucleotide distance threshold, ABGD, ASAP, and mPTP (Table 2, Table S5). A higher number of haplotypes correlated with lower efficiency (Table 2, Table S5).

The efficiency of four delimitation methods (*ad hoc* nucleotide distance threshold, ABGD, GMYC, and mPTP) was found to be positively related with the geographic distance between conspecific collection localities (Table 2, Table S5). For these methods a larger geographic range improved the proportion of matches.

GMYC resulted the only method to be affected by the taxonomic rank of the dataset (Table 2), with efficiency decreasing when genus rank datasets were analysed (Fig. 2B, Table S4, Table S5).

The presence in the dataset of species that are difficult to be identified (level II + level III) was found to decrease ABGD efficiency (Table 2, Table S5), with a higher proportion of merges related to difficult species. 3% threshold, *ad hoc* nucleotide distance threshold, ASAP, and mPTP were negatively affected by the presence of extremely difficult species (level III) (Table 2, Table S5), with a higher proportion of mixtures in extremely difficult species.

Discussion

Species delimitation

Our results suggested that the six molecular species delimitation methods considered in this study could be differentially affected by the dataset features, as expected considering their different assumptions and principles. Among the tested species delimitation methods, the highest efficiency resulted from distance-based method analyses (Fig. 2A), with *ad hoc* nucleotide distance threshold slightly outperforming ABGD and ASAP, and 3% threshold coming next. Our results confirm that ASAP achieves its aim of overcoming the two main limitations of ABGD, namely the need of defining a priori the maximum divergence of intraspecific diversity (P), and especially, the lack of scores associated to partitions, forcing ABGD users to choose independently the "best" partition among those resulting from the analysis (Puillandre et al., 2020). Interestingly, only in the 69% of the analysed datasets one of the two best ASAP-score partitions was identical to the partition selected as the "best" one from the ABGD output. The almost identical overall efficiency of the two methods notwithstanding the low proportion (69%) of shared selected partitions between them reveals that, regardless of the details of the methods, the result is comparable and ASAP may soon replace ABGD in the DNA taxonomy literature because of its advantages for the user.

In general, outperformance of distance-based methods over tree-based ones is consistent with what has been previously observed (Pentinsaari et al., 2017; Hoffman et al., 2019) when different species delimitation method performances were compared. In any case, it should be considered that the delimitation obtained in this study using distance-based methods (except for 3% nucleotide distance threshold) are not completely objective; in fact, the final choice of the best partition, among many potential choices for *ad hoc* nucleotide distance threshold and

ABGD, or only between two choices for ASAP, needs to be made by the user. Despite GMYC is known to be prone to oversplit species (Paz & Crawford, 2012; Hamilton, Hendrixson, Brewer, & Bond, 2014; Hoffman et al., 2019), in our analysis it was not associated with a high number of splits when compared to mPTP and the 3% threshold (Fig. 2); on the contrary, it underestimated species diversity, having the highest number of merged species. Despite COI intraspecific nucleotide distances of Chrysomelidae are known to be on average lower than 3% (Kubisz, Kajtoch, Mazur, & Rizun, 2011; Germain et al., 2013; Montagna et al., 2016), the species delimitation approach with such value as a fixed threshold led to a consistent number of splits. This result is due to the presence of some taxa with a high intraspecific variability, such as some species of the genus *Cassida*, which are known even to increase the optimal threshold for molecular identification of Cassidinae to 5.9% (Magoga et al., 2018). For this reason, 40% of Cassidine morphospecies were split using 3% threshold delimitation method. A similar situation was observed for Cryptocephalinae, where delimitation analysis using 3% threshold led to a consistent number of splits (22.5%, corresponding to 33 species out of 147). Despite the high proportion of closely related species groups characterized by low values of interspecific distances within this subfamily (Montagna, Sassi, & Giorgi, 2013; Montagna et al., 2016; Magoga et al., 2018) that strongly decrease the optimal threshold value for molecular identification (~1%, Magoga et al., 2018), values of intraspecific variability higher than the 5% were estimated for some Cryptocephalinae species, probably due to differences in term of evolutionary histories and/or ecology among the species (Magoga et al., 2018). Despite 3% not being in principle the most likely threshold value for separating intra- to inter-interspecific variability of the analysed taxa, the molecular delimitation using this fixed distance threshold had only a slightly lower efficiency than other distance-based methods (Fig. 2A).

Factors affecting species delimitation methods efficiency

One of the factors found to affect the majority of the species delimitation methods adopted in this study was the mean number of haplotypes per dataset. Higher mean numbers of haplotypes were significantly related to a decreasing efficiency of all the methods, except for GMYC (Table 2). This result could support the hypothesis that a higher number of sampled haplotypes correlates with a higher probability to find intermediate haplotypes among closely related species, in particular when sampling scale is large (Meyer & Pauly, 2005; Bergsten et al., 2012; Pentinsaari, Hebert, Mutanen, 2014; Phillips, Gillis, & Hanner, 2019). A wider intraspecific sampling could thus hide the barcoding gap and cause indecisive delimitations by adopting distance-based methods, consequently decreasing their efficiency (Fig. 3). Furthermore, several cases of splits were related to the 3% thresholds, which appeared too low to delimit the intraspecific level of the most haplotype-abundant species (Fig. 3). Contrary to what has been observed in this study, a positive relation between the number of haplotypes and the species delimitation efficiency was previously observed for coalescent-tree based methods (assuming an identical migration rate), with tendency to oversplit species when delimitation was implemented on datasets including a low number of haplotypes per species (Lohse, 2009). In our analyses, a low number of haplotypes per species did not result in an increase of the splits, neither for GMYC nor for mPTP; on the contrary, mPTP frequently split haplotypes-rich species. In fact, within the analysed datasets intraspecific sampling was often unbalanced, including a large number of equally and poorly sampled species and few oversampled species. Oversampled species resulted significantly prone to be oversplit by mPTP (Table S6), contrary to what was observed to occur using a previous version of the PTP method (Zhang et al., 2013). The efficiency of four methods (i.e., *ad hoc* nucleotide distance threshold, ABGD, GMYC, and mPTP) resulted to significantly increase when the median geographic distance among the

457 sampling localities of conspecifics increases. It is already known that, when intraspecific
458 sampling is incomplete, the main factor affecting the efficiency of coalescent tree-based
459 methods is the migration rate among species demes, given that high migration rate corresponds
460 to low delimitation efficiency and *vice versa* (Lohse, 2009). Our results are in agreement with
461 this evidence since the delimitation efficiency increases when the geographic distance among
462 conspecific collecting points increases, a condition that likely reduces or prevents the migration
463 of individuals among demes.
464 Despite that *ad hoc* nucleotide distance threshold and ABGD efficiency resulted good on the
465 majority of datasets, low values of median geographic distance among conspecifics sampling
466 localities were found to be related with more inaccurately delimited species. Different factors
467 could have affected the recovery of the barcoding gap in these cases. In some datasets where
468 geographic distance values were low and closely related species were included, the methods
469 tended to merge different species in the same partition (Fig. 4A). Possible explanations rely on
470 the recent divergence of the taxa and/or on the incomplete lineage sorting phenomenon, both
471 preventing the recovery of the barcoding gap when within-species geographic sampling is
472 limited. Higher efficiency on datasets with high median geographic distance among
473 conspecifics sampling localities values (e.g., *Agelastica* = 1,470 km; *Lilioceris* = 1,528 km;
474 *Plagiosterna* = 1,172 km) could be related to the species composition of those datasets. Since
475 the analysed data were generated by barcoding studies performed within four European areas,
476 the most common and widely distributed west-Palearctic species have likely been collected
477 from all the investigated areas. These species show low values of nucleotide intraspecific
478 distances indicating a panmictic-like condition that appears to promote delimitation efficiency
479 (Fig. 4B). Moreover, a sub-optimal efficiency has been observed for species-rich datasets,
480 where a clear barcoding gap occurs less frequently (Dellicour & Flot, 2018; Hoffman et al.,

2019), corresponding to intermediate values of median geographic distance among conspecifics sampling localities.

Further factors negatively affecting delimitation methods efficiency (except for GMYC) were the presence in the analysed datasets of species that are difficult and extremely difficult to be morphologically identified (difficulty level II+III and difficulty level III). A consistent behaviour between variability of the COI marker and species morphology is plausible; in the cases of recently diverged species, a high morphological similarity is mirrored by the lack of enough mutations on COI gene to make it effective for distinguishing between species (Nice & Shapiro, 1999; van Velzen, Weitschek, Felici, & Bakker, 2012; Chapple & Ritchie, 2013). On the other hand, despite the proven accuracy of morphological identification of specimens included in this study (Gómez-Rodríguez, Crampton-Platt, Timmermans, Baselga, & Vogler, 2014; Pentinsaari, Hebert, & Mutanen, 2014; Hendrich et al., 2015; Magoga et al., 2018), misidentifications may have implications in the observed pattern, especially in relation to species of extremely difficult identification.

GMYC resulted also the only method whose efficiency could be significantly affected by the taxonomic rank, in particular a decrease in the efficiency was observed with genus level data (Fig. 2B). In accordance to what was already observed by Talavera et al. (2013), the lack of significance of GMYC analyses in separating species within a genus could be related to the very low number of species included in some datasets, rather than to the taxonomic rank of the dataset itself. For our datasets, the majority of the not significant GMYC analyses (41 over 43) were indeed related to datasets including ≤ 3 species (39 genus datasets, 2 subfamily datasets). Even if the influence of the taxonomic rank on the other delimitation methods efficiency was not significant, it was possible to observe that an appreciably better efficiency of distance-based methods was present on genus datasets (Fig. 2B).

Conclusions

The results of this study demonstrated how 3% distance threshold, *ad hoc* nucleotide distance threshold, ABGD, ASAP, GMYC, and mPTP delimitation efficiency are influenced by different data-related factors including the number of haplotypes per species, the geographic distance between sampling points of individuals of the same species, the difficulty related to species morphological identification, and the taxonomic rank of the analysed dataset. Despite the difficulty of planning a species delimitation study accounting for all these factors, their impact on the delimitation should be considered and possibly avoided before performing any delimitation analysis. If their avoidance is not possible, they should be carefully considered as a caveat before drawing any conclusion from the results. Moreover, the quite common practice of comparing the results of various molecular species delimitation methods in species delimitation studies could also avoid to accept improper delimitations, considering that data-related factors differentially affected the efficiency of the methods.

Acknowledgements

This study was partially supported by MIUR-FFABR 2017/2018 provided to Matteo Montagna. Part of the study was developed within the PhD project of Giulia Magoga.

References

Ahrens, D., Fujisawa, T., Krammer, H. J., Eberle, J., Fabrizi, S., & Vogler, A. P. (2016). Rarity and Incomplete Sampling in DNA-Based Species Delimitation. *Systematic Biology*, *65*(3), 478–494. https://doi.org/10.1093/sysbio/syw002

Bergsten, J., Bilton, D. T., Fujisawa, T., Elliott, M., Monaghan, M. T., Balke, M., … Vogler, A. P. (2012). The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology*, *61*(5), 851–869. https://doi.org/10.1093/sysbio/sys037

Bezdek, J., & Baselga, A. (2015). Revision of western Palaearctic species of the *Oulema melanopus* group, with description of two new species from Europe (Coleoptera:

Chrysomelidae: Criocerinae). *Acta Entomologica Musei Nationalis Pragae*, *55*(1), 273–304.
Biondi, M., Urbani, F., & D'Alessandro, P. (2013). Endemism patterns in the Italian leaf beetle fauna (Coleoptera, Chrysomelidae). *ZooKeys*, *332*, 177–205. https://doi.org/10.3897/zookeys.332.5339
Birky, C. W. Jr, Adams, J., Gemmel, M., & Perry, J. (2010). Using Population Genetic Theory and DNA Sequences for Species Detection and Identification in Asexual Organisms. *PLoS One, 5*(5), e10609. https://doi.org/10.1371/journal.pone.0010609
Brown, S. D. J., Collins, R. A., Boyer, S., Lefort, M.-C., Malumbres-Olarte, J., Vink, C. J., & Cruickshank, R. H. (2012). SPIDER: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources*, *12*(3), 562–565. https://doi.org/10.1111/j.1755-0998.2011.03108.x
Brunetti, M., Magoga, G., Iannella, M., Biondi, M., & Montagna, M. (2019). Phylogeography and species distribution modelling of *Cryptocephalus barii* (Coleoptera: Chrysomelidae): is this alpine endemic species close to extinction? *ZooKeys*, *856*, 3–25. https://doi.org/10.3897/zookeys.856.32462
Burlini, M. (1955). Revisione dei *Cryptocephalus* italiani e della maggior parte delle specie di Europa (Col. Chrysomelidae). *Memorie della Società Entomologica Italiana*, *34*, 5–287.
Ceccarelli, F. S., Sharkey, M. J., & Zaldívar-Riverón, A. (2011). Species identification in the taxonomically neglected, highly diverse, neotropical parasitoid wasp genus *Notiospathius* (Braconidae: Doryctinae) based on an integrative molecular and morphological approach. *Molecular Phylogenetic and Evolution*, *62*(1), 485–495. https://doi.org/10.1016/j.ympev.2011.10.018
Chapple, D. G., & Ritchie, P. A. (2013). A Retrospective Approach to Testing the DNA Barcoding Method. *PLoS ONE*, *8*(11), e77882. https://doi.org/10.1371/journal.pone.0077882
Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*, *9*(8), 772. https://doi.org/10.1038/nmeth.2109
Davis, J. I., & Nixon, K. C. (1992). Populations, genetic variation, and the delimitation of phylogenetic species. *Systematic Biology*, *41*(4), 421–435. https://doi.org/10.2307/2992584
Dayrat, B. (2005). Towards integrative taxonomy. *Biological Journal of the Linnean Society*, *85*(3), 407–417. https://doi.org/10.1111/j.1095-8312.2005.00503.x
Dellicour, S., & Flot, J. (2018). The hitchhiker's guide to single-locus species delimitation. *Molecular ecology resources*, *18*(6), 1234–1246. https://doi.org /10.1111/1755-0998.12908
DeSalle, R., & Goldstein, P. (2019). Review and Interpretation of Trends in DNA Barcoding. *Frontiers in Ecology and Evolution*, *7*, 302. doi: 10.3389/fevo.2019.00302
deWaard, J. R., Levesque-Beaudin, V., deWaard, S. L., Ivanova, N. V., McKeown, J. T. A., Miskie, R., … Hebert, P. D. N. (2019). Expedited assessment of terrestrial arthropod diversity by coupling Malaise traps with DNA barcoding. *Genome*, *62*(3), 85–95. https://doi.org/10.1139/gen-2018-0093
Dincă, V., Montagud, S., Talavera, G., Hernández-Roldán, J., Munguira, M. L., García-Barros, … Vila, R. (2015) DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. *Scientific Reports*, *5*, 12395. https://doi.org/10.1038/srep12395
Doguet, S. (1994). *Coléoptères Chrysomelidae Volume 2 Alticinae – Faune de France, France*

et Regions Limitrophes. Paris, FR: Federation francaise des societes de sciences naturelles.
Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Phylogenetic and Evolution*, *29*(8), 1969–1973. https://doi.org/10.1093/molbev/mss075
Dupuis, J. R., Roe, A. D., & Sperling, F. A. (2012). Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Molecular Ecology*, *21*(18), 4422–4436. https://doi.org/10.1111/j.1365-294X.2012.05642.x
Esselstyn, J. A., Evans, B. J., Sedlock, J. L., Anwarali Khan, F. A., & Heaney, L. R. (2012). Single-locus species delimitation: A test of the mixed Yule–coalescent model, with an empirical application to Philippine round-leaf bats. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1743), 3678–3686. https://doi.org/10.1098/rspb.2012.0705
Ezard, T., Fujisawa, T., & Barraclough, T. G. (2009). splits: SPecies' LImits by Threshold Statistics. R package version 1.0.19. http://R-Forge.R-project.org/projects/splits/
Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biological Letters*, *4*(4), 423–425. https://doi.org/10.1098/rsbl.2008.0118
Flot, J., Couloux, A., & Tillier, S. (2010). Haplowebs as a graphical tool for delimiting species: a revival of Doyle's "field for recombination" approach and its application to the coral genus Pocillopora in Clipperton. *BMC Evolutionary Biology*, *10*, 372. https://doi.org/10.1186/1471-2148-10-372
Fox, J., & Weisberg, S. (2019). *An {R} Companion to Applied Regression, Third Edition*. Thousand Oaks, CA: Sage.
Fujisawa, T., & Barraclough, T. G. (2013). Delimiting species using single- locus data and the Generalized Mixed Yule Coalescent approach: A revised method and evaluation on simulated data sets. *Systematic Biology*, *62*(5), 707–724. https://doi.org/10.1093/sysbio/syt033
Fuss, G., Geiser, E. J., & Patzner, R. A. (2005). On the host plants of several leaf beetles of Central Europe - the problem of fame and evidence (Coleoptera: Chrysomelidae). *Koleopterologische Rundschau*, *75*, 359–371.
Gaytán, Á., Bergsten, J., Canelo, T., Pérez-Izquierdo, C., Santoro, M., & Bonal, R. (2020). DNA Barcoding and geographical scale effect: The problems of undersampling genetic diversity hotspots. *Ecology and Evolution*, *10*(19), 10754–10772. https://doi.org/10.1002/ece3.6733
García-Melo, J. E., Oliveira, C., Da Costa Silva, G. J., Ochoa-Orrego, L. E., Garcia Pereira, L. H., & Maldonado-Ocampo, J. A. (2019). Species delimitation of neotropical Characins (Stevardiinae): Implications for taxonomy of complex groups. *PLoS ONE*, *14*(6), e0216786. https://doi.org/10.1371/journal.pone.0216786
Germain, J., Chatot, C., Meusnier, I., Artige, E., Rasplus, J., & Cruaud, A. (2013). Molecular identification of *Epitrix* potato flea beetles (Coleoptera: Chrysomelidae) in Europe and North America. *Bulletin of Entomological Research*, *103*(4), 354–362. https://doi.org/10.1017/S000748531200079X
Göker, M., García-Blázquez, G., Voglmayr, H., Tellería, M. T., & Martín, M. P. (2009). Molecular Taxonomy of Phytopathogenic Fungi: A Case Study in Peronospora. *PLoS One*, *4*(7), e6319. https://doi.org/10.1371/journal.pone.0006319
Gómez-Rodríguez, C., Crampton-Platt, A., Timmermans, M. J. T. N., Baselga, A., & Vogler, A. P. (2014) Validating the power of mitochondrial metagenomics for community ecology and phylogenetics of complex assemblages. *Methods in Ecology and Evolution, 6*(8), 883–894. https://doi.org/10.1111/2041-210X.12376

Gómez-Zurita, J., Cardoso, A., Coronado, I., De la Cadena, G., Jurado-Rivera, J. A., Maes, J. M., … Papadopoulou, A. (2016). High-throughput biodiversity analysis: Rapid assessment of species richness and ecological interactions of Chrysomelidae (Coleoptera) in the tropics. *ZooKeys*, *597*, 3–26. https://doi.org/10.3897/zookeys.597.7065

Good, D. A., & Wake, D. B. (1992). Geographic variation and speciation in the torrent salamanders of the genus *Rhyacotriton* (Caudata: Rhyacotritonidae). *University of California Publication in Zoology*, *126*, 1–91.

Goodall-Copestake, W. P., Tarling, G. A., & Murphy, E. J. (2012). On the comparison of population-level estimates of haplotype and nucleotide diversity: a case study using the gene cox1 in animals. *Heredity*, *109*(1), 50–56. https://doi.org/10.1038/hdy.2012.12

Hallmann, C. A., Sorg, M., Jongejans, E., Henk, S., Hofland, N., Schwan, H., … Kroon, H. (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

Hamilton, C. A., Hendrixson, B. E., Brewer, M. S., & Bond, J. E. (2014). An evaluation of sampling effects on multiple DNA barcoding methods leads to an integrative approach for delimiting species: a case study of the North American tarantula genus *Aphonopelma* (Araneae, Mygalomorphae, Theraphosidae). *Molecular Phylogenetics and Evolution*, *71*(1), 79–93. https://doi.org/10.1016/j.ympev.2013.11.007

Hebert, P. D. N., Cywinska, A., Ball, S. L., & De Waard, J. R. (2003). Biological identifications through DNA barcodes. *Proceeding of the Royal Society, Biological Science, London*, *270*(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218

Hebert, P. D. N., & Gregory, T. R. (2005). The Promise of DNA Barcoding for Taxonomy. *Systematic Biology*, *54*(5), 852–859. https://doi.org/10.1080/10635150500354886

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*(4), 795–818. https://doi.org/10.1111/1755-0998.12354

Hickerson, M. J., Meyer, C. P., & Moritz, C. (2006). DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology*, *55*(5), 729–739. https://doi.org/10.1080/10635150600969898

Highton, R. (1989). Biochemical evolution in the slimy salamanders of the *Plethodon glutinosus* complex in the eastern United States. Part I. Geographic protein variation. Ill. *Illinois Biological Monographs*, *57*, 1–78.

Highton, R. (2000) Detecting Cryptic Species Using Allozyme Data. In R.C. Bruce, R. G. Jaeger, L. D. Houck L.D. (Eds.) *The Biology of Plethodontid Salamanders*. Boston, MA: Springer. https://doi.org/10.1007/978-1-4615-4255-1_10

Hijmans, R. J. (2017). Geosphere: Spherical Trigonometry. R package version 1.5-7. https://CRAN.R-project.org/package=geosphere.

Hofmann, E. P., Nicholson, K. E., Luque-Montes, I. R., Köhler, G., Cerrato-Mendoza, C. A., Medina-Flores, M., … Townsend, J. H. (2019). Cryptic Diversity, but to What Extent? Discordance Between Single-Locus Species Delimitation Methods Within Mainland Anoles (Squamata: Dactyloidae) of Northern Central America. *Frontiers in Genetics*, *10*, 11. https://doi.org/10.3389/fgene.2019.00011

Kajtoch, Ł., Montagna, M., & Wanat, M. (2018). Species delimitation within the *Bothryorrhynchapion* weevils: Multiple evidence from genetics, morphology and ecological associations. *Molecular Phylogenetics and Evolution*, *120*, 354–363. https://doi.org/10.1016/j.ympev.2017.12.022

Kapli, P., Lutteropp, S., Zhang, J., Kobert, K., Pavlidis, P., Stamatakis, A., & Flouri, T. (2017). Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics*, *33*(11), 1630–1638. https://doi.org/10.1093/bioinformatics/btx025

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*(2), 111–120. https://doi.org/10.1007/BF01731581

Kubisz, D., Kajtoch, Ł., Mazur, M., & Rizun, V. (2011). Molecular barcoding for central-eastern European *Crioceris* leaf-beetles (Coleoptera: Chrysomelidae). *Central European Journal of Biology*, *7*(1), 69–76. https://doi.org/10.2478/s11535-011-0099-4

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lohse, K. (2009). Can mtDNA barcodes be used to delimit species? A response to Pons et al. (2006). *Systematic Biology*, *58*(4), 439–442. https://doi.org/10.1093/sysbio/syp039

Magoga, G., Coral Sahin, D., Fontaneto, D., & Montagna, M. (2018). Barcoding of Chrysomelidae of Euro-Mediterranean area: efficiency and problematic species. *Scientific Reports*, *8*, 13398. https://doi.org/10.1038/s41598-018-31545-9

Mardulyn, P., Milinkovitch, M. C., & Pasteels, J. M. (1997). Phylogenetic Analyses of DNA and Allozyme Data Suggest that *Gonioctena* Leaf Beetles (Coleoptera; Chrysomelidae) Experienced Convergent Evolution in their History of Host-Plant Family Shifts. *Systematic Biology*, *46*(4), 722–747. https://doi.org/10.1093/sysbio/46.4.722

Mason, N. A., Fletcher, N. K., Gill, B. A., Funk, W. C., & Zamudio K. R. (2020). Coalescent-based species delimitation is sensitive to geographic sampling and isolation by distance. *Systematics and Biodiversity*, *18*(3), 269–280. https://doi.org/10.1080/14772000.2020.1730475

Mende, M., Biström, O., Meichssner, E., & Kölsch, G. (2010). The aquatic leaf beetle *Macroplea mutica* (Coleoptera: Chrysomelidae) in Europe: Population structure, postglacial colonization and the signature of passive dispersal. *European Journal of Entomology*, *107*(1), 101–113. https://doi.org/10.14411/eje.2010.013

Meyer, C. P., & Paulay, G. (2005). DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, *3*(12), e422. https://doi.org/10.1371/journal.pbio.0030422

Mioduchowska, M., Czyż, M. J., Goødyn, B., Kur, J., & Sell, J. (2018). Instances of erroneous DNA barcoding of metazoan invertebrates: Are universal cox1 gene primers too universal? *PLoS One*, *13*(6), e0199609. https://doi.org/10.1371/journal.pone.0199609

Monaghan, M. T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D. J. G., … Vogler, A. P. (2009). Accelerated Species Inventory on Madagascar Using Coalescent-Based Models of Species Delineation. *Systematic Biology*, *58*(3), 298–311. https://doi.org/10.1093/sysbio/syp027

Montagna, M., Sassi, D., & Giorgi, A. (2013). *Pachybrachis holerorum* (Coleoptera: Chrysomelidae: Cryptocephalinae), a new species from the Apennines, Italy, identified by integration of morphological and molecular data. *Zootaxa*, *3741*(2), 243–253. https://doi.org/10.11646/zootaxa.3741.2.3

Montagna, M., Kubisz, D., Mazur, M. A., Magoga, G., Scibior, R., & Kajtoch, L. (2016). Exploring species-level taxonomy in the *Cryptocephalus flavipes* species complex

(Coleoptera: Chrysomelidae). *Zoological Journal of Linnean Society, 179,* 92–109. https://doi.org/10.1111/zoj.12445
Montagna, M., Berruti, A., Bianciotto, V., Cremonesi, P., Giannico, R. Gusmeroli, F., … Gandini, G. (2018) Differential biodiversity responses between kingdoms (plants, fungi, bacteria and metazoa) along an Alpine succession gradient. *Molecular Ecology*, *27*(18), 3671–3685. https://doi-org.pros.lib.unimi.it:2050/10.1111/mec.14817
Müller, G. (1953). *I Coleotteri della Venezia Giulia. II. Coleoptera Phytophaga*. Trieste, IT: Centro Sperimentale Agrario e Forestale.
Mutanen, M., Kaila, L., & Tabell, K. (2013). Wide-ranging barcoding aids discovery of one-third increase of species richness in presumably well-investigated moths. *Scientific Reports, 3,* 2901. https://doi.org/10.1038/srep02901
Nadein, K. S. (2013), Febraina: a new subtribe of Alticini with cladistic analysis based on morphology (Coleoptera: Chrysomelidae: Galerucinae). *Systematic Entomology*, *38*(3), 491–506. https://doi.org/doi:10.1111/syen.12007
Nice, C. C., & Shapiro, A. M. (1999). Molecular and morphological divergence in the butterfly genus *Lycaeides* (Lepidoptera: Lycaenidae) in North America: evidence of recent speciation. *Journal of Evolutionary Biology*, *12,* 936–950. https://doi.org/10.1046/j.1420-9101.1999.00111.x
Nie, R.-e., Andújar, C., Gómez-Rodríguez, C., Bai, M., Xue, H.-J., Tang, M., … Vogler, A. P. (2020). The phylogeny of leaf beetles (Chrysomelidae) inferred from mitochondrial genomes. *Systematic Entomology*, *45*(1), 188–204. https://doi.org/10.1111/syen.12387
Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289–290. https://doi.org/10.1093/bioinformatics/btg412
Pavan-Kumar, A., Gireesh-Babu, P., & Lakra, W. S. (2015). DNA Metabarcoding: A New Approach for Rapid Biodiversity Assessment. *Journal of Cell Science and Molecular Biology*, *2*(1), 111.
Paz, A., & Crawford, A. J. (2012). Molecular-based rapid inventories of sympatric diversity: a comparison of DNA barcode clustering methods applied to geography-based vs clade-based sampling of amphibians. *Journal of Biosciences*, *37*(5), 887–896. https://doi.org/10.1007/s12038-012-9255-x
Pentinsaari, M., Hebert, P. D. N., Mutanen, M. (2014). Barcoding Beetles: A Regional Survey of 1872 Species Reveals High Identification Success and Unusually Deep Interspecific Divergences. *PLoS ONE, 9*(9): e108651. https://doi.org/10.1371/journal.pone.0108651
Pentinsaari, M., Vos, R., & Mutanen, M. (2017). Algorithmic single-locus species delimitation: effects of sampling effort, variation and nonmonophyly in four methods and 1870 species of beetles. *Molecular Ecology Resources*, *17*(3), 393–404. https://doi.org/10.1111/1755-0998.12557
Petitpierre, E. (2016). *Fauna ibérica: volume 13: Coleoptera: Chrysomelidae I*. Madrid, ES: Museo Nacional de Ciencias Naturales / Consejo Superior de Investigaciones Científicas.
Petitpierre, E. (2019). *Fauna ibérica: volume 46: Coleoptera: Chrysomelidae II*. Madrid, ES: Museo Nacional de Ciencias Naturales / Consejo Superior de Investigaciones Científicas.
Phillips, J. D., Gillis, D. J., & Hanner, R. H. (2019). Incomplete estimates of genetic diversity within species: Implications for DNA barcoding. *Ecology and Evolution*, *9*(5), 2996–3010. https://doi.org/10.1002/ece3.4757
Piper, R. W., & Compton, S. G. (2010). Population size and dispersal ability of *Cryptocephalus nitidulus* (Linnaeus, 1758) (Col.: Chrysomelidae). *Entomologist's Record and Journal of Variation*, *122*, 257–264.

773 Plewa, R., Sikora, K., Gutowski, J. M., Jaworski, T., Tarwacki, G., Tkaczyk, M., … Kajtoch,
774 Ł. (2018). Morphology, genetics and *Wolbachia* endosymbionts support distinctiveness of
775 *Monochamus sartor sartor* and *M. s. urussovii* (Coleoptera: Cerambycidae). *Arthropod*
776 *Systematics & Phylogeny*, *76*(1), 123–135.
777 Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., … Vogler,
778 A. P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed
779 insects. *Systematic Biology*, *55*(4), 595–609. https://doi.org/10.1080/10635150600852011
780 Porter, A. H. (1990). Testing nominal species boundaries using gene flow statistics: the
781 taxonomy of two hybridizing admiral butterflies (*Limenitis*: Nymphalidae). *Systematic*
782 *Zoology*, *39*(2), 131–147. https://doi.org/10.2307/2992451
783 Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, Automatic Barcode Gap
784 Discovery for primary species delimitation. *Molecular Ecology*, *21*(8), 1864–1877.
785 https://doi.org/10.1111/j.1365-294X.2011.05239.x
786 Puillandre, N., Brouillet, S., & Achaz, G. (2020). ASAP: assemble species by automatic
787 partitioning. *Molecular Ecology Resources*, *21*(2), 609–620. https://doi.org/10.1111/1755-
788 0998.13281
789 R Core Team (2019). R: A language and environment for statistical computing. R Foundation
790 for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
791 Reid, N. M., & Carstens, B. C. (2012). Phylogenetic estimation error can decrease the accuracy
792 species delimitation: a Bayesian implementation of the General Mixed Yule Coalescent
793 model. *BMC Evolutionary Biology*, *12*(1), 196. https://doi.org/10.1186/1471-2148-12-196
794 Revelle, W. (2018). psych: Procedures for Personality and Psychological Research. R package
795 version 1.9.12.31. https://CRAN.R-project.org/package=psych.
796 Rubinoff, D., & Holland, B. S. (2005). Between two extremes: mitochondrial DNA is neither
797 the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic Biology*,
798 *54*(6), 952–961. https://doi.org/10.1080/10635150500234674
799 Ruppert, K. M., Kline, R. J., & Rahman, M. S. (2019). Past, present, and future perspectives of
800 environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring,
801 and applications of global eDNA. *Global Ecology and Conservation*, *17*, e00547.
802 https://doi.org/10.1016/j.gecco.2019.e00547
803 Sánchez-Bayo, F., & Wyckhuys, K. A. G. (2019). Worldwide decline of the entomofauna: A
804 review of its drivers. *Biological Conservation*, *232*, 8–27.
805 https://doi.org/10.1016/j.biocon.2019.01.020
806 Schmitt, M., & Rönn, T. (2011). Types of geographical distribution of leaf beetles
807 (Chrysomelidae) in Central Europe. *ZooKeys*, *157*, 131–158.
808 https://doi.org/10.3897/zookeys.157.1798
809 Seifert, K. A. (2009). Progress towards DNA barcoding of fungi. *Molecular Ecology*
810 *Resources*, *9*(s1), 83–89. https://doi.org/10.1111/j.1755-0998.2009.02635.x
811 Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical Taxonomy*. San Francisco, CA: W. H.
812 Freeman and Company.
813 Strauss, S. Y. (1988). The Chrysomelidae: A useful group for investigating Herbivore-
814 Herbivore interactions. In P. Jolivet, E. Petitpierre, T. H. Hsiao (Eds.) *Biology of*
815 *Chrysomelidae. Series Entomologica vol 42* (pp. 91-105). Dordrecht, NL: Springer.
816 https://doi.org/10.1007/978-94-009-3105-3
817 Swartz, E. R., Mwale, M., & Hanner, R. (2008). A role for barcoding in the study of African
818 fish diversity and conservation. *South African Journal of Science*, *104*(7), 293–298.
819 Talavera, G., Dinca, V., & Vila, R. (2013). Factors affecting species delimitations with the
820 GMYC model: insights from a butterfly survey. *Methods in Ecology and Evolution*, *4*(12),

1101–1110. https://doi.org/10.1111/2041-210X.12107
Telfer, A. C., Young, M. R., Quinn, J., Perez, K., Sobel, C. N., Sones, J. E., … deWaard, J. R. (2015) Biodiversity inventories in high gear: DNA barcoding facilitates a rapid biotic survey of a temperate nature reserve. *Biodiversity data journal*, (3), e6313. https://doi.org/10.3897/BDJ.3.e6313
Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x
Tang, C. Q., Humphreys, A. M., Fontaneto, D., & Barraclough, T. G. (2014). Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. *Methods in Ecology and Evolution*, *5*(10), 1086–1094. https://doi.org/10.1111/2041-210X.12246
van Klink, R., Bowler, D. E., Gongalsky, K. B., Swengel, A. B., Gentile, A., & Chase, J. M. (2020). Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science*, *368*(6489), 417–420. https://doi.org/10.1126/science.aax9931
van Velzen, R., Weitschek, E., Felici, G., & Bakker, F. T. (2012). DNA Barcoding of Recently Diverged Species: Relative Performance of Matching Methods. *PLoS One*, *7*(1), e30490. https://doi.org/10.1371/journal.pone.0030490
Warchalowski, A. (2003). *Chrysomelidae. The leaf-beetles of Europe and the Mediterranean area*. Warszawa, PL: Nautura Optima Dux Fundation.
Yule, G. U. (1925). A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F. R. S. *Philosophical Transactions of the Royal Society of London Series B*, *213*(1925), 402–420.
Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, *29*(22), 2869–2876. https://doi.org/10.1093/bioinformatics/btt499

Data accessibility

All sequences analysed are publicly available in GenBank, their accession numbers are listed in Table S1, as well as the coordinates of collection localities. Information on each analysed dataset, from which GLMEM and GLMs independent variables are derived, is reported in Table S2. Results of each performed species delimitation analysis are summarized in Table S4. R scripts used to perform statistical analyses and to tabulate species delimitation results according to categories (match, merge, mixture, split) are available at https://github.com/MontagnaLab/species-delimitation-methods-comparison.

Author contributions

G.M., M.M. and D.F. designed the study. M.M. inferred species delimitation input trees. D.F. and G.M. performed the statistical analyses. G.M. analysed the data and wrote the manuscript. M.M. and D.F. contribute to the final version of the manuscript.

Table legends

Table 1. Effect of species delimitation method, number of haplotypes, geographic distance, taxonomic difficulty and taxonomic level on efficiency. The results are reported as a Type II analysis-of-variance table on a Generalised Linear Mixed Effects Model (GLMEM) object.

Table 2. Effect of number of haplotypes, geographic distance, taxonomic difficulty and taxonomic level on efficiency separately for each species delimitation method. The results are reported as a Type II analysis-of-variance table on Generalised Linear Model (GLM) objects.

Figure legends

Fig. 1. Rationale of the study. (a). datasets, molecular species delimitation analyses, and categorization of results. (b). statistical analyses.

Fig. 2. Efficiency of the species delimitation methods. (a). Mean percentage of match (red), merge (yellow), mixture (light blue), and split (purple) observed for each method. (b). Efficiency of the species delimitation methods recovered in three taxonomic levels, i.e., family (blue), subfamily (light blue) and genus (teal).

Fig. 3. Impact of the number of haplotypes on the 3% threshold delimitation efficiency. Distribution of pairwise nucleotide distances in five exemplar datasets (a-e), corresponding to different efficiency values and mean number of haplotypes; efficiency in percentage and the mean number of haplotypes per dataset are reported in the scatterplot as points coordinates. Regression line is a mere example of the relation between the variables.

Fig. 4. Impact of the geographic distance between conspecific collection localities on the ABGD and *ad hoc* threshold delimitation efficiency. The boxplot shows the distribution of median values of geographic distance between conspecific collection localities of all analysed datasets. Two datasets representatives for lower and higher geographic distance values (α and γ), which correspond to higher and lower species delimitation efficiency, are reported. (a) *Lochmaea* dataset consisting of four species; the map reports species collection localities. (b) *Plagiosterna* dataset consisting of one species; the map reports the collection localities of the different haplotypes recovered for *Plagiosterna aenea*. Minimum-spanning haplotypes networks show haplotypes relations (connecting lines) and abundance (size of the circles).

Supporting information

Table S1. COI sequences analysed in the study. Specimens taxonomy, GenBank accession

numbers and collection localities.
Table S2. Datasets analysed and explanatory variables.
Table S3. Nucleotide distance threshold values for species delimitation *ad hoc* estimated for each dataset.
Table S4. Molecular species delimitation analyses results.
Table S5. Output of GLMs related to each delimitation method.
Table S6. Relation between number of split resulting from mPTP delimitation and number of sequences per species analysed.

Fig. S1. Result of the test of multiple correlations between the explanatory variables. The diagonal reports the histograms of each variable; the scatterplots below the diagonal report the visual correlation between pairs of variables; the numbers in the squares above the diagonal report the Pearsons's correlation r values for each pair of variables.