UNIVERSITÀ DEGLI STUDI DI MILANO

DOTTORATO DI RICERCA IN
Agricoltura, Ambiente e Bioenergia XXXII Ciclo

Dipartimento di Scienze Agrarie e Ambientali – DiSAA

Molecular identification and delimitation of insect taxa: development of new data, approaches and evaluation of tools efficiency

GIULIA MAGOGA

Tutor
Dr. MATTEO MONTAGNA
Co-tutor
Dr. DIEGO FONTANETO

Coordinatore del Dottorato
prof. DANIELE BASSI

2019-2020

Index

1. Introduction and aim of the thesis 5

1.1 References 8

2. Results 12

2.1 Research article I: "Barcoding of Chrysomelidae of Euro-Mediterranean area: efficiency and problematic species" (Published in Scientific Reports 12

2.1.1 Summary 12

2.1.2 Manuscript 12

2.1.3 References 26

2.1.4 Appendix 29

2.1.5 Personal contribution to the work 29

2.2 Research article II: "DNA Barcoding as useful tool to identify crop pest flea beetles of Turkey" (Published in Journal of Applied Entomology) 30

2.2.1 Summary 30

2.2.2 Manuscript 30

2.2.3 References 45

2.2.4 Appendix 49

2.2.5 Personal contribution to the work 49

2.3 Research article III: "Morphology, genetics and *Wolbachia* endosymbionts support distinctiveness of *Monochamus sartor sartor* and *M. s. urussovii* (Coleoptera: Cerambycidae)" (Published in Arthropod Systematics & Phylogeny) 50

2.3.1 Summary 50

2.3.2 Manuscript 50

2.3.3 References 67

2.3.4 Appendix 72

2.3.5 Personal contribution to the work 72

2.4 Research article IV: "Molecular species delimitation of the Asian chestnut gall wasp biocontrol agent released in Italy" (Published in Insect Systematics and Evolution) 73

2.4.1 Summary 73

2.4.2 Manuscript 73

2.4.3 References 90

2.4.4 Personal contribution to the work 93

2.5 Research article V: “Factors affecting the efficiency of molecular species delimitation: evidences from the species rich beetle family of Chrysomelidae” (in preparation) 94
2.5.1 Summary 94
2.5.2 Manuscript 94
2.5.3 References 116
2.5.4 Personal contribution to the work 120

3. Conclusions 121

1 Introduction and aim of the thesis

Taxonomy is the science that classifies and identifies organism. Born from the human need of naming organisms, the traditional taxonomy is based on morphological characters whose differences are used for distinguishing taxa. Despite the unquestionable importance of morphological taxonomy, some limits related to its use are well known, e.g. the operator subjectivity in defining characters similarity, the long time needed for identifying certain organisms, the difficulty in the identification of juvenile stages or partial organisms, and, in particular in the last years, the decrease in the number of alfa-taxonomists (Rodman & Cody, 2003). In order to find a way to overcome morphological taxonomy limits, others approaches to taxonomy were looked for. After DNA discovery, taxonomists have taken advantages from information provided by DNA data for identifying and delimiting species, and in particular in the last decades, with the improving of molecular techniques, molecular taxonomy is become widely used.

For years, the most employed approach for molecularly identifying species has been DNA-barcoding (Hebert et al., 2003a), which relies on the idea that a nucleotide sequence can be used as unique barcode for distinguishing a species. The success of this identification method has not been the same for all taxonomic groups (Baker et al., 2009; Spooner, 2009; Kelly et al., 2011) since a universal marker that allows to distinguish all living organisms is not yet discovered. The application of this method is limited to the identification of the organisms for which a good genetic marker, enough informative for discriminating at species level, has been found. Barcoding is widely used for Metazoa, where most of the taxa can be identified using as marker a segment of the cytochrome oxidase subunit 1 gene (*cox1*, Hebert et al., 2003b). Starting from barcode idea, it is possible to reach the identification of an organism of unknown identity matching its barcode DNA sequence with those of identified organisms registered in databases. Since within species there is variability in the barcode sequences, and the databases are usually not so complete to include the whole variability of the species, this match could not be a perfect match. For this reason, the comparison among barcodes is based on nucleotide distances, and the more distant are two barcodes the higher is the probability that they belong to different species. Hence the main question of molecular identification through barcoding: "How similar two barcodes should be for being considered as belonging to the same species?". Focusing on *cox1*, many studies have been performed for finding a standard value of nucleotide distance that could be used as threshold between intra and interspecific nucleotide variability (Hebert et al., 2003a, Hebert et al., 2003b; Hebert et al., 2004). Despite some tentative values have been proposed (e.g. 3% of nucleotide divergence, Smith et al. 2005; the 10 times rule, Hebert et al. 2004), a universal threshold suitable

for all groups of organisms cannot be found, due to their different evolutionary histories; even if this it is routinely adopted in some groups of organisms, in particular bacteria and fungi. What can be done to overcome the highlighted limitation is to find the most suitable threshold for each organisms group analysing intra- and inter-specific nucleotide variability within it. As mentioned before, the correct molecular identification of an organism is strongly dependent also from the completeness, and equally important, from the accuracy of barcodes databases. For developing a dataset of barcode sequences, after specimens collection, each individual is subject to morphological species identification, subsequently to DNA-extraction and PCR targeting the barcode segment; once obtained the barcode nucleotide sequence, it is labelled with the name of the species it came from, and some test are performed for understanding if it is reliable for the identification of the species. One of the most adopted ways for testing the efficiency of a barcode dataset is to compare in pair all the de-novo produced sequences in order to check that it has the close best match, in term of nucleotide distance, with a sequences identified as belonging to the same species (Meier et al., 2006). Once the accuracy is assessed, the sequences can be registered in on-line databases and be available for barcode users. In the last years, a plethora of data have been developed and made available on-line in the major databases of *cox1* sequences (NCBI https://www.ncbi.nlm.nih.gov/; BOLD Ratnasingham & Hebert, 2007), leading to a strong improvement of molecular based identification of metazoans using this marker.

Within the field of molecular taxonomy, another innovative approach to taxonomy is the use of molecular information for the delimitation of species. Historically, molecular species delimitation was performed for testing the presence of gene flow between populations comparing alloenzyme variability (Porter, 1990; Good & Wake, 1992; Highton, 2000) or looking for the genotype profile of species (Davis & Nixon, 1992). Nowadays, the majority of molecular species delimitation studies exploit some bioinformatics algorithms, called species delimitation methods (Pons et al., 2006; Puillandre et al. 2012; Reid & Carstens, 2012; Zhang et al., 2013; Kapli et al., 2017) that allow researcher to infer hypothetic species starting from nucleotide sequences without any other information about the analysed taxa. Species delimitation methods are mainly used for delimiting species when others approaches are not effective (Lumley & Sperling 2010; Barley et al., 2013), for performing fast survey of the diversity of a geographic area (Barraclough et al., 2009; Gómez-Zurita et al., 2016) or as integration to other sources of information in new species descriptions (Birky et al., 2011; Ekimova et al., 2016; Scarpa et al., 2017) or for species groups taxonomic revisions (Stech et al., 2013; Montagna et al., 2016; Škaloud et al., 2016). Despite the use of information coming from multiple gene markers (nuclear and mitochondrial) has been demonstrated to improve molecular species delimitation (Rubinoff & Holland 2005; Dupuis et al., 2012), single-

marker species delimitations are currently frequent; in particular, species delimitation is often performed on *cox1* sequences (Dumas et al., 2015; Che et al., 2017), probably due to the efficiency of the marker in discriminating at species level and to the high number of sequences already available. For this reason, some of the most popular species delimitation methods have been developed or become commonly applied on *cox1* sequences (Puillandre et al. 2012; Kapli et al., 2017; Fujisawa and Barraclough, 2013). Among them there are some of the so called nucleotide-distance based methods and coalescent-tree based methods; the first approach of species delimitation starting from barcode sequences to be proposed was the use of a nucleotide distance threshold, a value of nucleotide divergence that can be used to discriminate between intraspecific and interspecific nucleotide variability (Hebert et al., 2003a; Meyer and Paulay, 2005). This method relies on the barcoding gap concept (Meyer and Paulay, 2005), i.e. the presence of a gap in the frequency distribution of nucleotide distances that is recognized as the point of transition between intra- and inter-specific level. Basing on barcoding gap concept, an algorithm for species delimitation named Automatic Barcode Gap Discovery was proposed (e.g. ABGD, Puillandre et al. 2012). ABGD automatically finds the distance at which the gap is located and consequently the threshold value for delimiting species; the benefit respect to the use of the previously mentioned distance threshold is that ABGD takes in account the possibility that multiple thresholds are present among taxa. Differently from the previous methods, coalescent-tree based methods account for the evolutionary relations among taxa in performing the delimitation. Starting from a dendrogram inferred from nucleotide sequences, these methods aim to find a threshold between intraspecific and interspecific branching basing on Yule/speciation and coalescent processes (Pons et al., 2006; Reid & Carstens, 2012; Zhang et al., 2013; Kapli et al., 2017). The usefulness of species delimitation methods is unquestioned, in particular thinking about how many taxonomic issues have been resolved thanks to their application, but it is also known that the results they generate can be sometimes influenced from features related to the analysed data. A bunch of studies on simulate datasets have been performed for testing the impact on species delimitation analyses of some factors, i.e. the effect of sampling scale, variation of the effective population size, of speciation rate and mutation rate, marker length, evolutionary history of species considered and sampling size (Lohse, 2009; Reid & Carstens, 2012; Esselstyn et al., 2012, Fujisawa & Barraclough, 2013; Dellicour & Flot 2018); since only some of these factors are derivable from real data and few studies were performed accounting for the influence of factors that can be planned in the study experimental design, more effort is needed for defining how to make better use of molecular species delimitation methods in practice.

Species identification and delimitation methods can be considered more useful for studying the taxa for which defining species limits is more complicated (Bruns et al., 2018); or that over time have been poorly studied (Ceccarelli et al., 2011). Within Metazoa, one of such groups is insects; due to the high diversity of insects, the taxa present in some geographic areas are almost unknown or the taxonomy of others is conflictual. Within insects, also in groups that can be considered well studied, often further investigations lead to new species discovery or taxonomic incongruences highlighted. In this thesis, molecular taxonomy methods are applied and tested for efficiency on insects, with a special focus on leaf beetles (Coleoptera, Chrysomelidae). Chrysomelidae is a species-rich family of phytophagous Coleoptera that includes taxa with different ecology, both in term of exploited host plants and specialization level (monophagy, oligophagy, polyphagy). Beside the importance of this group in the study of biodiversity, it has also an economically relevant interest related to those species that are pests of crops or ornamental plants (Kaygin et al., 2008; Zhao et al., 2008; Sawadogo et al., 2015). Moreover, Chrysomelidae represents an excellent group for investigating insect-host plant relation (Wang et al., 2018; Yıldız et al., 2019) or the evolution of trophic specialization (Gikonyo et al., 2019; Salvi et al., 2019). Both applied and basic research can benefit in exploring the taxonomy of these beetles using molecular tools.

In detail, in my thesis DNA-barcoding molecular identification approach is applied for developing *cox1* barcode libraries for Euro-Mediterranean Chrysomelidae identification and DNA-barcoding tested for its efficiency in the identification of the species of the family. Moreover, the aspect of the use of nucleotide distance thresholds for species molecular identification is better investigated, developing and comparing specific thresholds for Chrysomelidae taxa in view of understanding the error related to the use of a fixed threshold. The second part of this thesis consists in the application of molecular species delimitation methods to other insects taxa with the aim of resolving their taxonomic status in an integrative taxonomy framework. Finally, starting from concerns and issue derived from the application of molecular species delimitation methods to different datasets, partially related to the planning of an adequate experimental design and on the other hand also to the biology of the analysed species, the influence of some dataset-related factors on species delimitation efficiency was evaluated.

1.1 References

Baker, A. J., Tavares, E. S. & Elbourne, R. F. (2009) Countering criticisms of single mitochondrial DNA gene barcoding in birds. *Molecular Ecology Resources* **9**: 257-268.

Barley, A. J., White, J., Diesmos, A. C. & Brown, R. M. (2013) The challenge of species delimitation at the extremes: diversification without morphological change in Philippine sun skinks. *Evolution* **67**: 3556-3572.

Barraclough, T. G., Hughes, M., Ashford-Hodges, N. & Fujisawa, T. (2009) Inferring evolutionarily significant units of bacterial diversity from broad environmental surveys of single-locus data. *Biology letters* **5**: 425-428.

Birky, C. W., Ricci, C., Melone, G. & Fontaneto, D. (2011) Integrating DNA and morphological taxonomy to describe diversity in poorly studied microscopic animals: new species of the genus Abrochtha Bryce, 1910 (Rotifera: Bdelloidea: Philodinavidae). *Zoological Journal of the Linnean Society* **161**: 723-734.

Bruns, T. D., Corradi, N., Redecker, D., Taylor, J. W. & Öpik, M. (2018) Glomeromycotina: what is a species and why should we care?. *New Phytologist* **220**: 963-967.

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: 485-495.

Che, Y., Gui, S., Lo, N., Ritchie, A. & Wang, Z. (2017) Species Delimitation and Phylogenetic Relationships in Ectobiid Cockroaches (Dictyoptera, Blattodea) from China. *PLoS ONE* **12**: e0169006.

Davis, J. I. & Nixon, K. C. (1992) Populations, genetic variation, and the delimitation of phylogenetic species. *Systematic Biology* **41**: 421-435.

Dellicour, S. & Flot, J. (2018) The hitchhiker's guide to single-locus species delimitation. *Molecular ecology resources* **18:** 1234-1246.

Dumas, P., Barbut, J., Le Ru, B., Silvain, J. F., Clamens, A. L., d'Alençon, E. & Kergoat, G. J. (2015) Phylogenetic Molecular Species Delimitations Unravel Potential New Species in the Pest Genus *Spodoptera* Guenée, 1852 (Lepidoptera, Noctuidae). *PLoS ONE* **10**: e0122407.

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**: 4422-4436.

Ekimova, I. A., Schepetov, D. M., Chichvarkhina, O. V. & Chichvarkhin, A. (2016) Nudibranch molluscs of the genus Dendronotus Alder et Hancock, 1845 (Heterobranchia: Dendronotina) from Northwestern Sea of Japan with description of a new species. *Invertebrate Zoology* **13**: 15-42.

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:** 3678-3686.

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**: 707-724.

Gikonyo, M. W., Biondi, M. & Beran, F. (2019) Adaptation of flea beetles to Brassicaceae: host plant associations and geographic distribution of *Psylliodes* Latreille and *Phyllotreta* Chevrolat (Coleoptera, Chrysomelidae). In: Schmitt, M., Chaboo, C. S., Biondi, M. (Eds) Research on Chrysomelidae 8. *ZooKeys* **856**: 51-73.

Gómez-Zurita, J., Cardoso, A., Coronado, I., De la Cadena, G., Jurado-Rivera, J. A., Maes, J. M., Montelongo, T., Nguyen, D. & 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.

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.

Hebert, P. D. N., Cywinska, A., Ball, S. L. & deWaard, J. R. (2003a) Biological identifications through DNA barcodes. *Royal Society* **270**: 313-321.

Hebert, P. D., Ratnasingham, S. & deWaard, J. R. (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings. *Biological sciences* **270**: 96-99.

Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S. & Francis, C. M. (2004) Identification of birds through DNA barcodes, *PLoS Biology* **2**: e312.

Highton, R. (2000) Detecting cryptic species using allozyme data. In The Biology of Plethodontid Luo, A., Ling, C., Ho, S. Y. W., Zhu C.-D. Comparison of Methods for Molecular Species Delimitation Across a Range of Speciation Scenarios. *Systematic Biology* **67:** 830-846.

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**: 1630-1638.

Kaygin, A. T., Sönmezyildiz, H., Ulgentürk, S. & Ozdemir, I. (2008) Insect species damage on ornamental plants and saplings of Bartin Province and its vicinity in the Western Black Sea region of Turkey. *International journal of molecular sciences* **9**: 526-541.

Kelly, L. J., Hollingsworth, P. M., Coppins, B. J., Ellis, C. J., Harrold, P., Tosh, J. & Yahr, R. (2011) DNA barcoding of lichenized fungi demonstrates high identification success in a floristic context. *New Phytologist* **191**: 288-300.

Lohse, K. (2009) Can mtDNA barcodes be used to delimit species? A response to Pons et al. (2006). *Systematic Biology* **8**: 439-442.

Lumley, L. M. & Sperling, F. A. (2010) Integrating morphology and mitochondrial DNA for species delimitation within the spruce budworm (Choristoneura fumiferana) cryptic species complex (Lepidoptera: Tortricidae). *Systematic Entomology* **35**: 416-428.

Meyer, C. P. & Paulay, G. (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology* **3**: 2229-2238.

Meier, R., Shiyang, K., Vaidya, G. & Ng, P. (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology* **55**: 715-728.

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.

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D. & Vogler, A. P. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* **55**: 595-609.

Porter, A. H. (1990) Testing nominal species boundaries using gene flow statistics: the taxonomy of two hybridizing admiral butterflies (*Limenitis*: Nymphalidae). *Systematic Zoology* **39**: 131-147.

Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G. (2012) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* **21**: 1864-1877.

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

Reid, N. M. & Carstens, B. C. (2012) Phylogenetic estimation error can decrease the accuracy species delimitation: a Bayesian implementation of the General Mixed Yule Coalescent model. *BMC Evolutionary Biology* **12**: 196.

Rodman, J. E. & Cody, J. H. (2003) The taxonomic impediment overcome: NSF's partnerships for enhancing expertise in taxonomy (PEET) as a model. *Systematic Biology* **52**: 428-435.

Rubinoff, D. & Holland, B. S. (2005) Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic Biology* **54**: 952-961.

Salvi, D., D'Alessandro, P. & Biondi, M. (2019) Host plant associations in Western Palaearctic *Longitarsus* flea beetles (Chrysomelidae, Galerucinae, Alticini): a preliminary phylogenetic assessment. In: Schmitt, M., Chaboo, C. S., Biondi, M. (Eds) Research on Chrysomelidae 8. *ZooKeys* **856**: 101-114.

Sawadogo, A., Nagalo, E., Nacro, S., Rouamba, M. & Kenis, M. (2015). Population Dynamics of *Aphthona whitfieldi* (Coleoptera: Chrysomelidae), Pest of *Jatropha curcas*, and Environmental Factors Favoring Its Abundance in Burkina Faso. *Journal of insect science* **15**: 108.

Scarpa, F., Cossu, P., Sanna, D., Lai, T., Casu, M. & Curini-Galletti, M. (2017) New insights on the genus *Otoplana* Du Plessis, 1889 (Platyhelminthes: Proseriata), with description of two new species from the Canary Islands. *Marine Biodiversity*. 1-13.

Smith, M. A., Fisher, B. L. & Hebert, P. D. N. (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philosophical Transactions of the Royal Society of London, Series B* **360**: 1825-1834.

Škaloud, P., Friedl, T., Hallmann, C., Beck, A. & Dal Grande, F. (2016) Taxonomic revision and species delimitation of coccoid green algae currently assigned to the genus *Dictyochloropsis* (Trebouxiophyceae, Chlorophyta). *Journal of Phycology* **52**: 599-617.

Spooner, D. M. (2009) DNA barcoding will frequently fail in complicated groups: An example in wild potatoes. *American Journal of Botany* **96**: 1177-1189.

Stech, M., Veldman, S., Larraín, J., Muñoz, J., Quandt, D., Hassel, K. & Kruijer, H. (2013) Molecular Species Delimitation in the *Racomitrium canescens* Complex (Grimmiaceae) and Implications for DNA Barcoding of Species Complexes in Mosses. *PLoS ONE* **8**: e53134.

Wang, D., Pentzold, S., Kunert, M., Groth, M., Brandt, W., Pasteels, J. M., Boland, W. & Burse, A. (2018) A subset of chemosensory genes differs between two populations of a specialized leaf beetle after host plant shift. *Ecology and Evolution* **8**: 8055- 8075.

Yıldız, D., Altun, N. & Bilgener, M. (2019) The Effect of Nutrient-Allelochemicals Interaction on Food Consumption and Growth Performance of Alder Leaf Beetle, *Agelastica alni* L. (Coleoptera: Chrysomelidae). *International Journal of Secondary Metabolite* **6**: 196-204.

Zhang, J., Kapli, P., Pavlidis, P. & Stamatakis, A. (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* **29**: 2869-2876.

Zhao Y., Yang, G., Wang-Pruski, G. & You, M. (2008) *Phyllotreta striolata* (Coleoptera: Chrysomelidae): Arginine kinase cloning and RNAi-based pest control. *European Journal of Entomology* **105**: 815-822.

2 Results

2.1 Research article

Barcoding of Chrysomelidae of Euro-Mediterranean area: efficiency and problematic species

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

2.1.1 Summary

Leaf beetles (Coleoptera: Chrysomelidae), with more than 37,000 species worldwide and about 2,300 in the Euro-Mediterranean region, are an ecological and economical relevant family, making their molecular identification of interest also in agriculture. This study, part of the Mediterranean Chrysomelidae Barcoding project (www.c-bar.org), aims to: (i) develop a reference Cytochrome c oxidase I (*cox1*) library for the molecular identification of the Euro-Mediterranean Chrysomelidae; (ii) test the efficiency of DNA barcoding for leaf beetles identification; (iii) develop and compare optimal thresholds for distance-based identifications estimated at family and subfamily level, minimizing false positives and false negatives. Within this study, 889 *cox1* nucleotide sequences of 261 species were provided; after the inclusion of information from other sources, a dataset of 7,237 sequences (542 species) was analysed. The average intra-interspecific distances were in the range of those recorded for Coleoptera: 1.6–24%. The estimated barcoding efficiency (~94%) confirmed the usefulness of this tool for Chrysomelidae identification. The few cases of failure were recorded for closely related species (e.g., *Cryptocephalus marginellus* superspecies, *Cryptocephalus violaceus* - *Cryptocephalus duplicatus* and some *Altica* species), even with morphologically different species sharing the same *cox1* haplotype. Different optimal thresholds were achieved for the tested taxonomic levels, confirming that group-specific thresholds significantly improve molecular identifications.

2.1.2 Manuscript

Introduction

Chrysomelidae, or leaf beetles, is one of the most species-rich families of Coleoptera. Leaf beetles are distributed worldwide (except Antarctica) and inhabit almost all habitats presenting vegetation. Leaf beetles include more than 37,000 species at global level belonging to more than 2,000 genera

(Jolivet et al., 2011). In the Palearctic region approximately 3,500 species have been described so far (Konstantinov et al., 2009), and about 2,300 of them occur in the Euro-Mediterranean region (Magoga et al., 2016; Blondel et al., 2010; Warchalowski, 2003). With few exceptions, leaf beetles are phytophagous insects adapted to feed on plant species, including some of agricultural interest (e.g., *Diabrotica virgifera* LeConte, 1868 and *Leptinotarsa decemlineata* Say, 1824). The species-specific strict association with the host plants makes leaf beetles an interesting group for evolutionary studies (e.g. Futuyma & McCafferty, 1990; Chung et al., 2017) however, they have received attention also due to their impact on agriculture (e.g. Sawadogo et al., 2015) and to their use as biological control agents of invasive plants (Grevstad, 2006; Szűcs et al., 2012). The correct identification of organisms is regarded to be essential in both applicative field and evolutionary studies; yet, their taxonomy, based on morphological features, requires a high level of expertise that could be reached only after years of study. In some cases, as for several species groups, the accurate species identification of adults can only be achieved extracting genitalia (Montagna et al., 2013; Sassi, 2014; Montagna et al., 2016). Therefore, preimaginal stages can be only rarely identified to species level. Thus, approaches based on morphology may not be efficient for beetles identification and become strongly time consuming especially in large scale studies, for example in biomonitoring surveys for agricultural biocontrol.

DNA based approaches have emerged as useful tools for the identification of organisms (Hebert et al., 2003; Hebert & Gregory, 2005) and the efficacy of the cytochrome oxidase subunit I (*cox1*) marker in molecular identification of Coleoptera (including leaf beatles) was demonstrated (García-Robledo et al., 2013; Lopes et al., 2015; Thormann et al., 2016). At present, on-line databases harbour about 8,000 *cox1* sequences assigned to approximately 1200 leaf beetles species worldwide, roughly about 4% of the overall described species. We are still far from having a reliable reference database, and most of the detailed barcoding studies for European leaf beetles have been developed within limited geographic contexts (Hendrich et al., 2015; Pentinsaari et al., 2014). In order to increase the number of barcoded species, including also the rare ones, large scale biodiversity studies focused on leaf beetles inhabiting different biogeographic regions are needed. The Mediterranean Chrysomelidae Barcoding project (C-bar, Magoga et al., 2016), started in 2009 and, involving many taxonomists and specialists of different subfamilies, aims to develop a reference database of sequences for the molecular identification of leaf beetles inhabiting the Euro-Mediterranean region. In the present study, we analyse the dataset of *cox1* gene sequences obtained within the C-bar project with the purpose of: i) evaluate the efficiency of the DNA barcoding dataset; ii) estimate the optimal intraspecific and interspecific thresholds for the identification of leaf beetles species at different taxonomic level (i.e., family *vs* subfamily level).

Materials and methods

Ethics Statement

No species of Coleoptera Chrysomelidae are listed in national laws as protected or endangered. All the specimens were collected between 2009-2013 in state-owned properties. The collection of these invertebrates is not subjected to restriction by national or international laws and does not require special permission. All the organisms were collected before the approval of Nagoya protocol 283/2014/UE.

Sample collection and identification

Leaf beetles were collected in sampling campaigns occurred from 2009 to 2013 in different ecoregions of central and southern Europe and North Africa. The animals were collected using different methods: from the vegetation by sweep net or by beating sheet, and directly by hand in specific habitats. All the specimens were stored in absolute ethanol in order to preserve the genomic DNA and preserved at -20°C. Specimen manipulation and dissection (when necessary) were completed with the auxiliary use of a stereomicroscope Leica MS5, a compound microscope Zeiss Axio Zoom V16, and images were acquired with the digital camera Zeiss Axiocam 506. The specimens were morphologically identified by expert taxonomists. The nomenclature adopted in this study follows that of the European Fauna (https://fauna-eu.org/).

DNA extraction, sequencing

DNA extractions were performed in two laboratories (the Biodiversity Institute of Ontario - University of Guelph; the Laboratory of Molecular Entomology at Dipartimento di Scienze Agrarie e Ambientali - Università degli Studi di Milano) adopting the following different protocols: i) DNA extraction from one hind leg of the specimen, and ii) DNA extraction from the whole body, in both cases using Qiagen DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) as reported in Magoga et al., 2016.

After DNA extraction, the voucher specimens were dry mounted on pins in the case of whole body DNA extraction, or preserved in absolute ethanol at -30°C in the case of DNA extraction from a single leg. An aliquot of the extracted DNA was preserved in both laboratories at -80°C as reference. The standard barcode region of the mitochondrial *cox1* was amplified by PCR using standard barcode primers LCO1490/HCO2198 (Folmer, 1994). In case of unsuccessful amplifications, the alternative *cox1* primers LepF1 / LepR1 were adopted to amplify the selected region (Hebert, 2004). PCRs were performed in a volume of 25 μL reaction mix containing: 1X

GoTaq reaction Buffer (10 mM Tris-HCl at pH 8.3, 50 mM KCl and 1.5 mM $MgCl_2$), 0.2 mM of each deoxynucleoside triphosphate, 0.5 pmol of each primer, 0.3 U of GoTaq DNA Polymerase and 10/20 ng of template DNA. The adopted thermal protocol is reported in Montagna et al., 2013. Positive amplicons were directly sequenced on both strands using the marker-specific primers from ABI technology (Applied Biosystems, Foster City, USA). Consensus sequences were obtained editing electropherograms using Geneious R8 (Biomatters Ltd., Auckland, New Zealand. License owned by Matteo Montagna). Spurious amplifications of *cox1* sequences were checked using Standard Nucleotide BLAST (Altschul et al., 1990). The presence of open reading frame was verified for the obtained sequences by using the on-line tool EMBOSS Transeq (http://www.ebi.ac.uk/Tools/st/emboss_transeq/), then sequences were aligned at codon level using MUSCLE (Edgar, 2004) in MEGA 6.06 (Tamura et al., 2013). Consensus sequences were deposited in the Bold Systems (Ratnasingham, 2007) and in the European Nucleotide Archive to make them available for future studies (accession numbers reported in Tab. S1).

Sequence mining and dataset development

Accession numbers of orthologous sequences belonging to European and Mediterranean Leaf Beatles species were recovered from previously published DNA barcoding studies (Hendrich et al., 2015; Gómez-Rodríguez et al., 2015; Pentinsaari et al., 2014) and used to download the corresponding nucleotide sequences from public repositories (i.e., BOLD and GeneBank); this operation was completed using the R 3.3.3 (R Core Team, 2017) library *ape* v4.1 (Popescu et al., 2012) and *rentrez* v1.1.0 (Winter, 2017).

Overall a total of 6,348 *cox1* gene sequences were retrieved from public repositories. These nucleotide sequences and those obtained in the present study were organised in two datasets: *i*) dataset *DS1*, composed only by the nucleotide sequences developed in this study; *ii*) dataset *DS2*, composed by the sequences mined from online databases plus dataset *DS1*. We keep separated the two datasets in order to evaluate the efficiency of the here developed dataset, and to estimate the barcoding efficiency for the whole family using all available *cox1* sequences (*DS2*).

Taxonomy was standardised checking for the presence of synonymous names and assigning only one name (the accepted one following European Fauna https://fauna-eu.org/ nomenclature) to each species. *DS1* and *DS2* were also split in sub-datasets in order to obtain datasets including only one leaf beetles subfamily each. Only subfamily level datasets consisting of at least 2 species were retained. The procedure led to obtain datasets for the following ten subfamilies: Alticinae, Cassidinae, Chrysomelinae, Criocerinae, Cryptocephalinae, Donaciinae, Eumolpinae, Galerucinae, Hispinae and Orsodacninae.

Bioinformatic analyses

For all the morphologically identified species of *DS1* and *DS2*, intraspecific and interspecific nucleotide divergences were calculated starting from a pairwise distance matrix developed using R library *spider* v1.1-5 (Brown et al., 2012) adopting Kimura-two parameters as substitution model (K2P) (Kimura, 1980). With the same R package a Threshold optimisation analysis was performed on *DS1*, *DS2* and on each subfamily-level dataset in order to calculate the value of nucleotide distance (optimal threshold; OT) that minimises the error related to molecular identification. This error is caused by the discordance between morphological and molecular identification and is called cumulative error (CE), calculated as the sum of the number of false positives (FP, conspecifics with a value of nucleotide divergence higher than the threshold value) plus the number of false negatives (FN, heterospecifics with a value of nucleotide divergence lower than the threshold value) (Meyer & Paulay, 2005) Differences in CE values estimated at family and subfamily level were assessed using Student *t* test. The efficiency of molecular identification was estimated performing Best Close Match analyses, defined by Meier et al. 2006, on *DS1* and *DS2* (family level). The method compares each sequence of dataset with the others included in it and checks if the best matches (i.e., pairs of sequences with the lowest values of nucleotide distance) are between sequences of organisms morphologically identified as the same species. Each best match results in one of the following four states: "correct", when the two closest sequences under the defined threshold belong to the same species; "incorrect", the opposite situation; "ambiguous", when the closest match is represented by more than one species; and, "no id" when no match is recorded under the chosen threshold.

For some groups of closely related species, where several misidentifications were observed, minimum-spanning haplotype networks (Bandelt et al., 1999) were reconstructed using PopArt (Leigh, 2015).

Data availability

All the *cox1* sequences and the metadata associated with the organisms processed in this study are available in Bold Systems, European Nucleotide Archive and Supplementary Table S1 and S3. Voucher specimens are deposited into M.M. private collection.

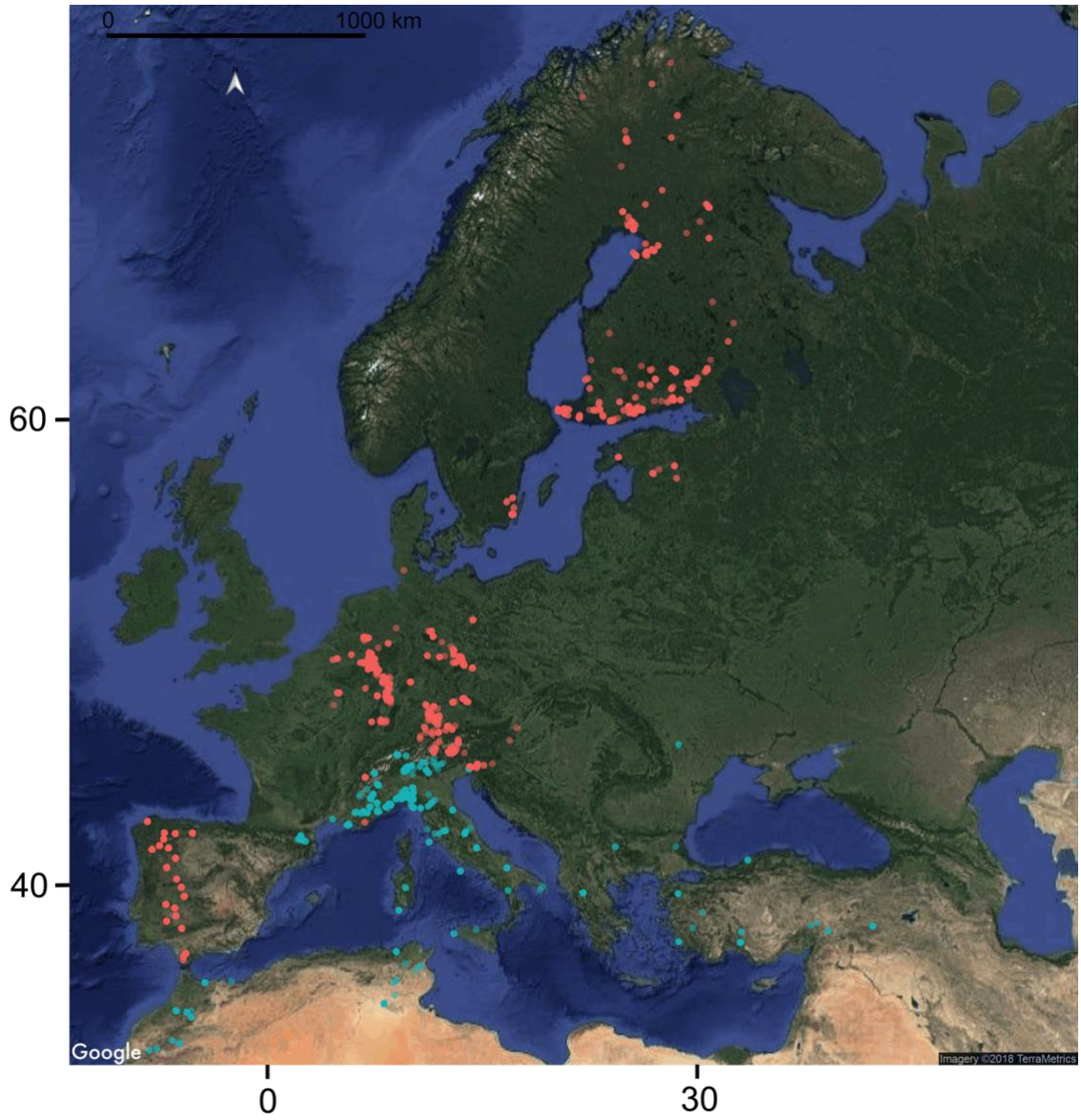

Figure 1. Collection sites of the individuals analysed in this study. Sampling localities of the individuals processed in this study (light blue dots) and whose barcodes were mined from online databases (orange dots). Map developed using R libraries *ggmap* (Kahle & Wickham, 2013), *ggplot2* (Wickham, 2009), *ggsn* (Baquero, 2017) background image downloaded using the cited libraries from Google Imagery©2018 TerraMetrics.

Results

General features

The dataset developed in this study (i.e., *DS1*) consists of 889 *cox1* sequences (average 654 bp [range: 494-658]), with a base composition of A = 29.4%; C = 19.6%; G = 16.1%; T = 34.9%. The dataset includes sequences of 261 leaf beetles species, the 11.4% of the Euro-Mediterranean species (74 singletons), belonging to 64 genera collected from ten countries within the Euro-Mediterranean

region (Fig. 1; Table S1). Out of the 261 barcoded species, *cox1* sequences of 52 species were not already present in any online repository (Table S4).

Dataset *DS2*, consisting of the previously available sequences with the addition of *DS1*, is composed by 7,237 *cox1* sequences (average 652 bp [range: 460-658]); with a base composition of A=29.6%; C=18.7%; G=16.1%; T=35.5%. In *DS2* the *cox1* sequences of 542 species (~24% of the Euro-Mediterranean fauna) sampled in 19 different countries of Europe and North Africa are included (Fig. 1; Table S5).

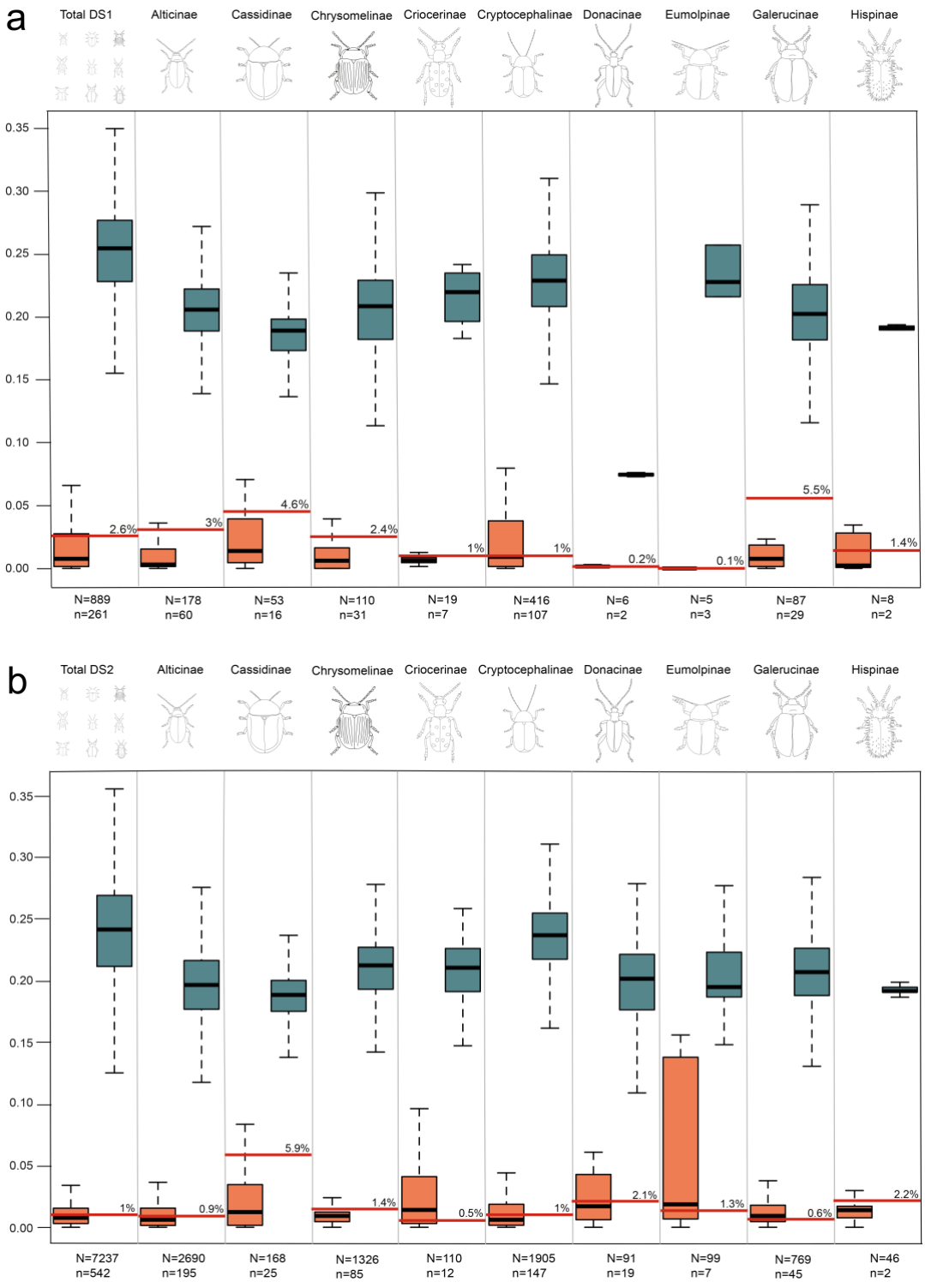

Figure 2. Boxplots of K2P inter-intraspecific pairwise nucleotide distances inferred from DS1 (a) and DS2 (b) datasets. Estimated intraspecific (orange) and interspecific (cadet blue) nucleotide distances are reported for each dataset at family and subfamily levels; optimal thresholds are reported as percentage and indicated by the red horizontal lines; below each bar the number of sequences (N) and of species (n) are reported. Above the bars datasets identifiers are reported.

Morphospecies intra-interspecific nucleotide distance

The distributions of intraspecific and interspecific pairwise nucleotide distances overlap, thus

resulting in the absence of a clear barcode gap in both family-level datasets (Fig. S1). The mean intraspecific nucleotide distance, estimated with the K2P nucleotide substitution model, resulted in 2% [0-20.6] for dataset *DS1* and 1.6% [0-27.6] for *DS2* (Fig. S1). The exceptionally high maximum value of the intraspecific nucleotide distance in *DS1* of 20.6% is the result of the comparisons between sequences of two *Lachnaia tristigma* (Lacordaire, 1848) populations, both collected in France in the Alpes-de-Haute-Provence department. The interspecific nucleotide distance resulted in 25.1% [0-37.1%] in the case of *DS1* and of 24% [0-43.2%] in the case of *DS2* (Fig. S1). Noteworthy, 0 or close to 0 values of nucleotide distance were recovered between specimens belonging to different species (Fig. S1); among others, exemplar cases are represented by: *Cryptocephalus violaceus* Laicharting, 1781 - *Cryptocephalus duplicatus* Suffrian, 1845; *Lachnaia italica* Weise, 1881 - *Lachnaia tristigma*; members of *Cryptocephalus marginellus* Olivier, 1791 species complex, and of the *Cryptocephalus hypochaeridis* (Linnaeus 1758) species complex. In detail, specimens of *C. duplicatus* collected in Turkey and of *C. violaceus* collected in Greece possessed the same *cox1* haplotype; within the *Cryptocephalus marginellus* complex, notable is the case of *Cryptocephalus renatae* Sassi, 2001 collected in Savona province having only ~0.6% of nucleotide distance from *C. marginellus* collected in a geographically close locality (Nice, FR) and of the *Cryptocephalus eridani* Sassi, 2001 having ~0.4% from *Cryptocephalus hennigi* Sassi, 2011, both collected in Cuneo province. The *cox1* haplotype network of *C. marginellus* complex (Fig. 3A) confirms the previous results and shows that the currently known species in the group are not well distinguished as species clusters; the only exceptions are represented by *Cryptocephalus zoiai* Sassi, 2001 and *Cryptocephalus aquitanus* Sassi, 2001, unambiguously separated from the other species (Fig. 3A). In addition, within the subfamily of Alticinae some specimens belonging to the following eight out of 14 *Altica* morphospecies present in the *DS2* showed the same or highly similar *cox1* haplotype (range of nucleotide intraspecific distances: 0-12.6% and of nucleotide interspecific distances: 0-13.8%): *Altica aenescens* (Weise, 1888), *Altica ampelophaga* Guérin-Meneville, 1858, *Altica ericeti* (Allard, 1859), *Altica brevicollis* Foudras, 1860, *Altica engstroemi* (Sahlberg, 1894), *Altica lythri* Aubé, 1843, *Altica longicollis* (Allard, 1860) and *Altica oleracea* (Linnaéus, 1758) (Fig. 3B).

Table 1. Cumulative error values related to optimal thresholds of *DS1* and *DS2* datasets at family and subfamilies level.

Dataset IDs	**Chr.**	**Alt.**	**Cas.**	**Chrys.**	**Cri.**	**Cry.**	**Don.**	**Eum.**	**Gal.**	**Ors.**	**His.**	**t test**	**p-value**
DS1	97	10	1	3	1	66	2	0	1	0	0	-13.7	<0.001
DS2	816	413	1	40	16	192	12	0	73	0	0	-17.6	<0.001

Abbreviations: Chr.: Chrysomelidae; Alt.: Alticinae; Cas.: Cassidinae; Chrys.: Chrysomelinae; Cri.: Criocerinae; Cry.: Cryptocephalinae; Don.: Donacinae; Eum.: Eumolpinae; Gal.: Galerucinae; Ors.:

Orsodacninae; His.: Hispinae.

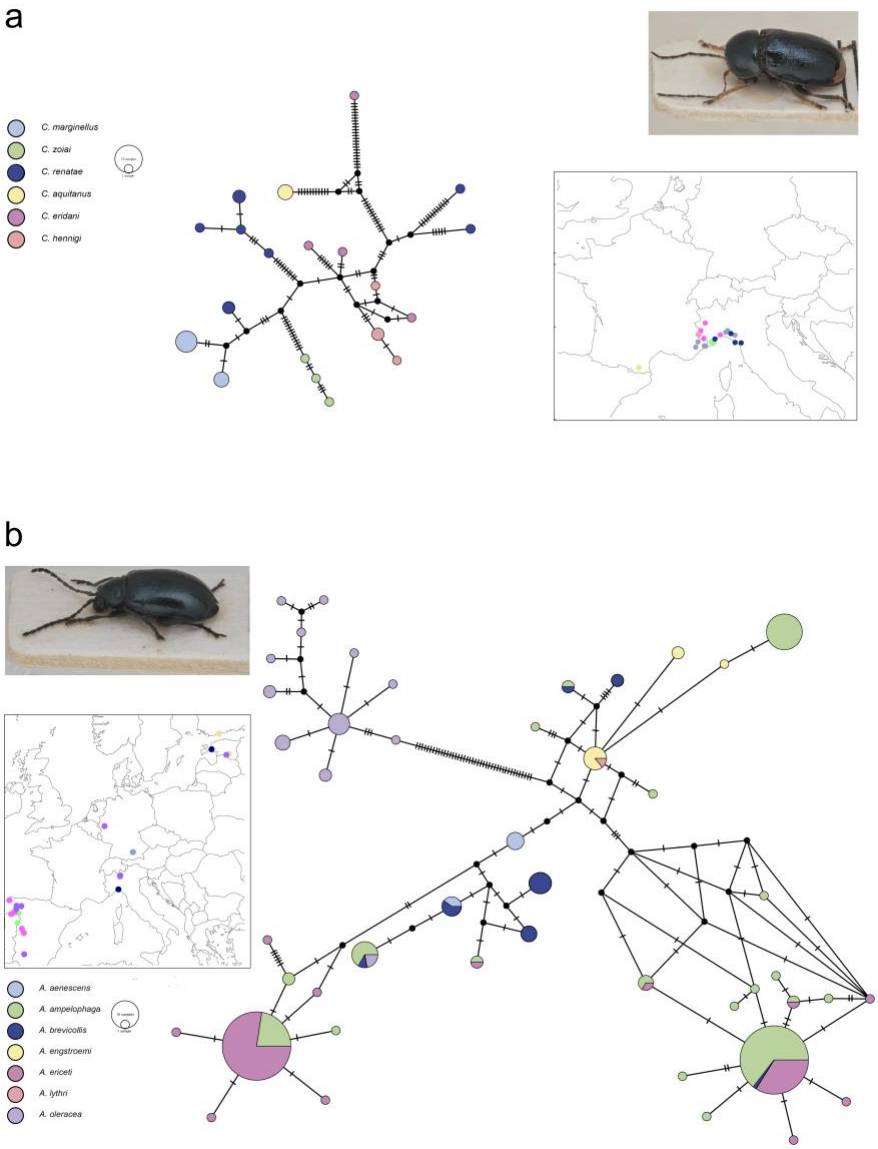

Figure 3. Minimum-spanning haplotype networks of *cox1* sequences. (a) *Cryptocephalus marginellus* superspecies. (b) *Altica oleracea* species complex. For each group is reported an image of the representative species (*C. marginellus* and *A. oleracea*, respectively) and a map reporting collecting sites of the specimens included in this study. Diameter of the circle is proportional to haplotypes abundance.

Optimal threshold and barcode efficiency

The optimal threshold that minimises the number of false positive (FP) and false negative (FN) identifications resulted in 2.6% of distance for *DS1*, with an associated cumulative error of 97 sequences out of 889 (10.9 %, FP=38, FN=59); for *DS2* it resulted in a value of 1% of nucleotide

distance, with a cumulative error of 816 sequences out of 7237, 11.3%, FP= 209, FN= 607). The sum of the cumulative errors obtained from optimal threshold analyses performed on the subfamily level datasets obtained from *DS1* resulted in 85 sequences (9.6%), and in 748 sequences (10.3%) in the case of datasets from *DS2*. (Table 1). These error values are significantly different from the cumulative errors obtained for the total datasets, i.e., *DS1*: t = -13.7, p-value < 0.001; *DS2*: t = -17.6, p-value < 0.001 (Table 1). The highest error values, related with subfamily datasets, were observed for Cryptocephalinae obtained from *DS1* (66 sequences out of 416, 15.8%, threshold 1%) and for Alticinae from *DS2* (413 out of 2690 sequences, 15.3%, threshold 0.9%). By contrast, the lowest error of only one sequence was obtained for both datasets of Cassidinae, with 53 sequences in *DS1* and 168 sequences in *DS2*; the associated OTs were higher than those observed for the other subfamilies, 4.6% for *DS1* and 5.9% for *DS2*.

The barcode efficiency of *DS1*, evaluated through the best close match analysis gave an *OT* of 2.6%, resulting in 93% of correct identification (828 out of 889); 58 species, consisting of a single *cox1* sequence, were considered as correctly identified since no match with other heterospecific sequences occurred. Of the 61 sequences that revealed identification errors, 14 were classified as incorrect identifications. These sequences belong to taxa between which very low interspecific nucleotide distances were observed (e.g. *C. hennigi* - *C. eridani*; *A. brevicollis* - *A. lythri*; *L. italica* - *L. tristigma*); in addition to these, they include also sequences from *Longitarsus apicalis* (Beck, 1817), showing the best match with *Longitarsus aeneicollis* (Faldermann, 1837) (pairwise nucleotide distance of 0.2%), and one specimen of *Oulema melanopus* (Linnaeus, 1758) that matched with *Oulema duftschmidi* (Redtenbacher, 1874) (1.2% of pairwise nucleotide distance). A total of 39 sequences (34 morphospecies) resulted in no match with conspecifics because of a pairwise nucleotide distance higher than the adopted OT. Among these cases, a sequence of *Cassida denticollis* Suffrian, 1844 showed about 15% of nucleotide divergence from other sequences assigned to the same species. The eight ambiguous identifications involve the sister species *C. violaceus*-*C. duplicatus* and *L. tristigma*.

The same analysis performed on *DS2* highlighted the presence of 94.1% correct identifications (6,811 sequences out of 7,237), 52 incorrect, 164 ambiguous and 210 missing identifications (Tab. S2), with an OT of 1%. Among incorrect and ambiguous identifications, beyond the *DS1* cases mentioned above, only one match involved at least one sequence from *DS1* (i.e., *Psylliodes brisouti* Bedel, 1898 MS0000647 with *Psylliodes instabilis* Foudras, 1860 KM445439). Incorrect and ambiguous identifications were observed also among the retrieved sequences: e.g. one sequence of *L. tristigma* and one of *Lachnaia gallaeca* Baselga & Ruiz-García, 2007 (nucleotide distance 0.2%); *Plateumaris sericea* (Linnaeus, 1761) and *Plateumaris discolor* (Panzer, 1795) sequences

and, *Galerucella pusilla* (Duftschmid, 1825) and *Galerucella calmariensis* (Linnaeus, 1767). As regard the missing identifications, sequences assigned to 11 species of *Cassida*, 27 of *Cryptocephalus* and 6 of *Smaragdina* genera did not match those of conspecifics because of intraspecific genetic distances higher than the OT.

Discussion

Identification efficiency

The results achieved by the performed analyses confirmed the usefulness of the DNA barcoding approach as a tool for the molecular identification of Chrysomelidae. The obtained identification efficiencies are comparable for both datasets; our dataset (N_{DS1}= 889 sequences) showed 93% of correct identifications, while 94% of correct identifications was obtained for *DS2* (i.e., the available *cox1* sequences + *DS1*; N_{DS2}= 7,237 sequences), which cover the ~24% of the Euro-Mediterranean species. The barcoding efficiency recovered in the present study is similar to those achieved in other studies dealing with beetles, as example 89% in the case of *Bembidion* species (Raupach et al., 2016), approximately 92% in the case of the Central European Coleoptera (39% of the fauna; Hendrich et al., 2015) and 100% in the case of *Crioceris* species (Kubisz et al., 2012). In any case, in these studies different approaches were adopted to estimate the barcoding efficiency, thus a direct comparison could not be performed.

Incorrect, ambiguous and missing identifications observed in our study are possibly related with the inability of DNA barcoding in identifying taxa in the presence of: *i*) superspecies (two or more close related species with allopatric distribution that can occasionally hybridise (Mayr, 1931)) and cryptic species complexes (Van Velzen et al., 2012; Jiang et al., 2014) ; *ii*) cases of hybridisation or introgression; *iii*) incomplete lineage sorting; and *iv*) bacterial endosymbionts changing pathways of mtDNA inheritance (Smith et al., 2012; Klopfstein et al., 2016). In these cases, the lack of a clear barcode gap between intraspecific and interspecific nucleotide distances vanish the possibility to identify species (Meyer & Paulay, 2005; Wiemers & Fiedler, 2007) The phenomenon is evident also in the analysed datasets, where a clear barcode gap cannot be found (Fig. S1).

Interestingly, the estimated optimal threshold of *DS2* was lower than that of *DS1*, 1% and 2.6% respectively. These results could be related to the different haplotype diversity and to the different taxonomic composition of the two datasets. The mean number of haplotypes per species of the two datasets is 6.7 (on average 13.4 sequences per species) and 2.5 (on average 3.4 sequences per species) in the case of *DS2* and *DS1*, respectively; thus, *DS1* possesses fewer sequences per species but a higher number of haplotype per species (approximately one haplotype per sequence) than *DS2*. The differences between the two datasets might be related to the sampling strategies adopted

in C-Bar project, where attempts have been made to maximise the number of conspecifics from different localities, rather than to process numerous specimens of the same species from the same locality.

Threshold optimisation analyses showed also a significant decrease of the cumulative error when OTs were estimated at the subfamily level in comparison to when they were estimated at the family level (Tab 1). Phylogenetically closely related species are supposed to have similar rates of nucleotide substitution due to shared morphological, biological and ecological traits (*e.g.*, number of generation per year, tendency to isolation of the populations due to the habitat structure or to the dispersal ability of the species (Fujisawa et al., 2015) and for this reason should be easier to define a reliable threshold between intraspecific and interspecific divergence. We can hypothesise that not all Chrysomelidae share the same rate in nucleotide substitutions, since different subfamilies are characterised by different morphological, ecological and physiological adaptation, as the Maulik's organ that confers jumping capabilities to Alticinae (Scherer, 1971; Furth, 1988), the limited dispersal capabilities of Chrysomelinae and Cryptocephalinae (Piper & Compton, 2003) or the presence of bacterial endosymbiont that, in the case of Donacinae, allows the larvae to survive in anoxic conditions under water (Kleinschmidt & Kölsch, 2011). Moreover, the different OTs achieved for Chrysomelidae subfamilies underline that the use of a unique threshold for the entire family decreases the identification efficiency of DNA barcoding (Table 1). Beyond classical barcoding studies, the implementation of group specific thresholds, leading to a more accurate taxonomic identification, should be also evaluated for OTUs clustering in metabarcoding analyses instead of the employment of fixed thresholds (as in the case of Fonseca et al., 2017; Potter et al., 2017).

Concerning the cumulative error, the highest value was obtained for Cryptocephalinae subfamily (*DS1*). This dataset, accounting for 46.8% of *DS1* sequences, includes different species complexes (e.g., *Cryptocephalus marginellus* and *Cryptocephalus hypochoeridis*). The presence of species complexes increases the overlap between intra and interspecific distances and consequently the cumulative error at the optimal threshold. In the case of *DS2*, Alticinae resulted the subfamily with the highest error associated to the OT of 0.9%. This finding could be associated to a high proportion of sequences belonging to the genus *Altica* in this dataset (229 out of 2,690), a taxon for which inconsistences between molecular and morphological signals were already found (Jäckel et al., 2013).

Molecular identification of closely related species

Barcode sequences of closely related species within the groups *Cryptocephalus hypochaeridis* (Leonardi & Sassi, 2001; Gómez-Zurita et al., 2011) and *Oulema melanopus* (Berti, 1989; Bezdek J. & Baselga, 2015) were here analysed; as expected, low values of nucleotide interspecific distances within groups were observed. Moreover, our study highlighted other interesting cases of sequences belonging to morphologically similar species groups not properly identified by best close match analyses. *Cryptocephalus marginellus* superspecies, including six species that differ in their distributions and in the shape of the median lobe of aedeagus (Sassi, 2001; 2011) represents one of these cases. These species are present in Spain (*C. aquitanus*), France (*C. aquitanus, C. marginellus, C. eridani, C. zoiai*), Italy (*C. eridani, C. marginellus, C. renatae, C. hennigi, C. zoiai)* and Switzerland (*C. eridani*), and their distributions partially overlap in some areas. The close relationships among these species highlighted by morphological features were here confirmed by the *cox1* variability and by the structure of the haplotype network (Fig. 3A); however, no shared haplotypes between species were observed. Well-separated clusters were recovered for *C. aquitanus* and *C. zoiai* that, in addition to *C. marginellus*, resulted the only monophyletic taxon within this complex (Fig. 3A). The analysis of pairwise nucleotide distances showed low values between different species, the lowest one between specimens collected in the area where the range of the species overlap (e.g., *C. eridani* - *C. hennigi*). Incomplete lineage sorting could be considered an explanation for these results, even if introgression between species with overlapping distribution has to be taken into account. A further interesting result concerns *C. violaceus* and *C. duplicatus*, two morphologically very similar species distinguishable only on the basis of the shape of the median lobe of the aedeagus. *C. violaceus* is present in central and southern Europe while *C. duplicatus* in the southern east of Europe and the Middle East. No nucleotide differences were observed between the *cox1* sequences of *C. violaceus* collected in Greece and *C. duplicatus* collected in Turkey. Since the distribution of the species overlaps in Greece, we can hypothesise recent events of introgression. This phenomenon is known to occur when, after an allopatric speciation, two sister species come in contact and establish an area of secondary sympatry; due to the lack of reproductive isolation they have the possibility to hybridise with the result of a stable integration of genetic material from one species into the other one (Mallet, 2005; Baack & Rieseberg, 2007).

Shared haplotypes were observed among the following *Altica* species: *A. ericeti-A. ampelophaga; A. ampelophaga-A. oleracea-A. brevicollis; A. brevicollis-A. aenescens; A. ericeti-A. ampelophaga-A. brevicollis; A. ampelophaga-A. brevicollis; A. lythri-A. engstroemi*. Identification of many species belonging to *Altica*, included those above mentioned, is not easy adopting morphological criterion; it is mainly based on the observation of adult male genitalia, which in some cases is not

totally informative because of the presence of intraspecific morphological variation (Aslan et al., 2004). In addition, adult females are often indeterminable (Warchalowski, 2010). This difficulty in species identification is also mirrored at the molecular basis, where the species of this group are unidentifiable using cox1 (Fig. 3B) as well as by using other mtDNA markers (Warchalowski, 2010). Morphological and cox1 nucleotide similarity suggests the possible need of a taxonomic revision of the *Altica* species mentioned above. The obtained results, *viz* the low interspecific nucleotide divergence and the presence of shared haplotypes, is congruent with a scenario of incomplete lineage sorting due to the recent origin of the group of species and hybridization. A further possibility, supported by the presence of different strains of the maternally inherited endosymbiont *Wolbachia* within and between *Altica* species, consists of a rapid spread within populations of ancestral or introgressed haplotypes, caused by the cytoplasmic incompatibility induced by *Wolbachia* (Jäckel et al., 2013). In this last scenario, *Wolbachia* might have played a crucial role in mating isolation and thus in the speciation process, as suggested for other groups of close related taxa (e.g., Jaenike et al., 2006; Plewa et al., 2018). Further studies, using genomic approaches, are required to disentangle among the reported possibilities.

Conclusion

This study provides cox1 sequences of 261 Chrysomelidae species (~12% of the Euro-Mediterranean Fauna; 889 barcodes) collected in the Euro-Mediterranean area (52 species new to on-line repositories) and confirms the usefulness and efficiency of DNA barcoding for the identification of these beetles. Cases of barcoding failure in identifying members of the family were observed especially for closely related species, and some of them are reported for the first time in this study. The comparisons among optimal thresholds estimated at different taxonomic levels, *viz* family and subfamily, have underlined the importance of using taxon-specific thresholds to increase the efficacy of molecular identification.

2.1.3 References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Aslan, I., Calmasur, O. & Bilgin, O. C. (2004) A morphometric study of *Altica oleracea* (Linnaeus, 1758) and *A. deserticola* (Weise, 1889) (Coleoptera: Chrysomelidae: Alticinae). *Entomologica Fennica* **15**: 1-5.

Baack, E. J. & Rieseberg, L. H. (2007) A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics & Development* **17**: 513-518.

Bandelt, H. J., Forster, P. & Röhl, A. (1999) Median-Joining Networks for Inferring Intraspecific Phylogenies. *Molecular Biology and Evolution* **16**: 37-48.

Baquero, O. S. (2017) North Symbols and Scale Bars for Maps Created with 'ggplot2' or 'ggmap' https://rdrr.io/cran/ggsn/

Berti, N. (1989) Contribution à la Faune de France. L'identité d'*Oulema* (O.) *melanopus* (L.) (Col. Chrysomelidae Criocerinae). *Bulletin de la Société entomologique de France* **94**: 47-57.

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**: 273-304.

Blondel, J., Aronson, J., Bodiou, J. Y. & Boeuf, G. (2010) The Mediterranean Region – Biological diversity in space and time. Oxford Univ. Press.

Brown, S. D., 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**: 562-565.

Chung, S. H., Scully, E. D., Peiffer, M., Geib, S. M., Rosa, C., Hoover, K. & Felton, G. W. (2017) Host plant species determines symbiotic bacterial community mediating suppression of plant defenses. *Scientific Reports* **7**: 39690.

Edgar, R. C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113.

Folmer, O., Black M., Hoeh W., Lutz, R. & Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**: 294-299.

Fonseca, V. G., Sinniger, F., Gaspar, J. M., Quince, C., Creer, S., Power, D. M., Peck, L. S. & Clark, M. S. (2017). Revealing higher than expected meiofaunal diversity in Antarctic sediments: a metabarcoding approach. *Scientific Reports* **7**: 6094.

Furth, D. G. (1988) The jumping apparatus of flea beetles (Alticinae) — The metafemoral spring in *Biology of Chrysomelidae* Eds. Jolivet P., Petitpierre E., Hsiao T. H. Springer.

Futuyma, D. J. & McCafferty, S. S. (1990) Phylogeny and the evolution of host plant association in the leaf beetle genus *Ophraella* (Coleoptera, Chrysomelidae). *Evolution* **44**: 1885.

Fujisawa, T., Vogler, A. P. & Barraclough, T. G. (2015) 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* **282**: 2014-2476.

García-Robledo, C., Kuprewicz, E. K., Staines, C. L., Kress, W. J. & Erwin, T. L. (2013) Using a comprehensive DNA barcode library to detect novel egg and larval host plant associations in a *Cephaloleia* rolled-leaf beetle (Coleoptera: Chrysomelidae). *Biological Journal of the Linnean society* **110**: 189-198.

Gómez-Zurita, J., Sassi, D., Cardoso, A. & Balke, M. (2011) Evolution of Cryptocephalus leaf beetles related to *C. sericeus* (Coleoptera: Chrysomelidae) and the role of hybridisation in generating species mtDNA paraphyly. *Zoologica Scripta* **41**: 47-67.

Gómez-Rodríguez, C., Crampton-Platt, A., Timmermans, M. J. T. N., Baselga, A. & Vogler, A. P. (2015) Validating the power of mitochondrial metagenomics for community ecology and phylogenetics of complex assemblages. *Methods in Ecology and Evolution* **6**: 883-894.

Grevstad, F. S. (2006) Ten-year impacts of the biological control agents *Galerucella pusilla* and *G. calmariensis* (Coleoptera: Chrysomelidae) on purple loosestrife (*Lythrum salicaria*) in Central New York State. *Journal of Biological Control.* **39**: 1-8.

Hebert, P. D. N., Cywinska, A., Ball, S. L. & De Waard, J. R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences* **270**: 313-321.

Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S. & Francis, C. M. (2004) Identification of birds through DNA barcodes. *PLoS Biol*ogy **2**. e312.

Hebert, P. D. N. & Gregory, T. R. (2005) The Promise of DNA Barcoding for Taxonomy. *Systematic Biology* **54**: 852-859.

Hendrich, L., Morinière, J., Haszprunar, G., Hebert, P. D., Hausmann, A., Köhler, F. & Balke, M. (2015) Acomprehensive 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.

Jäckel, R., Mora, D. & Dobler, S. (2013) Evidence for selective sweeps by *Wolbachia* infections: phylogeny of *Altica* leaf beetles and their reproductive parasites. *Molecular Ecology* **22**: 4241-4255.

Jaenike, J., Dyer, K. A., Cornish, C. & Minhas, M. S. (2006) Asymmetrical Reinforcement and *Wolbachia* Infection in *Drosophila*. *PLoS Biology* **4**: e325.

Jiang, F., Jin, Q., Liang, L., Zhang, A. B. & Li, Z. H. (2014) Existence of species complex largely reduced barcoding success for invasive species of Tephritidae: a case study in *Bactrocera* spp. *Molecular Ecology Resources.* **14**: 1114-1128.

Jolivet, P., Santiago-Blay, J. A. & Schmitt, M. (2011) Research on Chrysomelidae **3.** Pensoft Publishers.

Leigh, J. W. & Bryant, D. (2015) Popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6**: 1110-1116.

Leonardi, C. & Sassi, D. (2001) Studio critico sulle specie di *Cryptocephalus* del gruppo *hypochaeridis* (Linné, 1758) e sulle forme ad esse attribuite (Coleoptera, Chrysomelidae). *Atti della Società italiana di scienze naturali e del Museo civico di storia naturale di Milano.* **142**: 3-96.

Lopes, S. T., Dourado, C. G., Oliveira, A. R., Silva, I. F., Farinha, A., Rebelo, M. T. & Dias, D. (2015) Molecular Identification of Western-Palearctic Leaf Beetles (Coleoptera, Chrysomelidae). *Journal of the Entomological Research Society* **17**: 93-101.

Kahle, D. & Wickham, H. (2013) ggmap: Spatial Visualization with ggplot2. *R Journal* **5**: 144-16

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.

Kleinschmidt, B. & Kölsch, G. (2011) Adopting Bacteria in Order to Adapt to Water-How Reed Beetles Colonized the Wetlands (Coleoptera, Chrysomelidae, Donaciinae) *Insects* **9**: 540-54.

Klopfstein, S., Kropf, C. & Baur, H. (2016) *Wolbachia* endosymbionts distort DNA barcoding in the parasitoid wasp genus *Diplazon* (Hymenoptera: Ichneumonidae). *Zoological Journal of the Linnean Society* **177**: 541-557.

Konstantinov, A. S., Korotyaev, B. A. & Volkovitsh, M. G. (2009) Insect Biodiversity in the Palearctic Region in *Insect Biodiversity: Science and Society*. Ed. Foottit, R. G. and Adler, P. H. Wiley-Blackwell

Kubisz, D., Kajtoch, Ł., Mazur, M. A. & Rizun V. (2012) Molecular barcoding for central-eastern European Crioceris leaf-beetles (Coleoptera: Chrysomelidae). *Central European Journal of Biology* **7**: 69.

Magoga, G., Sassi, D., Daccordi, M., Leonardi, C., Mirzaei, M., Regalin, R., Lozzia, G. & Montagna, M. (2016) Barcoding Chrysomelidae: a resource for taxonomy and biodiversity conservation in the Mediterranean Region. In: Jolivet P, Santiago-Blay J, Schmitt M (Eds) Research on Chrysomelidae 6. *ZooKeys* **597**: 27-38.

Mallet, J. (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* **20**: 229-237.

Mayr, E. (1931) Birds collected during the Whitney South Sea Expedition. XII. Notes on *Halcyon chloris* and some of its subspecies. *American Museum Novitates* 469.

Meyer, C. P. & Paulay, G. (2005). DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology* **3:** 2229-2238.

Meier, R., Shiyang, K., Vaidya, G. & Ng, P. (2006). DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology* **55**: 715-728.

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**: 243-253.

Montagna, M., Kubisz, D., Mazur, M. A., Magoga, G., Ścibior, R. & Kajtoch, Ł. (2016) Exploring species-level taxonomy in the *Cryptocephalus flavipes* species complex (Coleoptera: Chrysomelidae). *Zoological Journal of Linnean Society* **179**: 92-109.

Pentinsaari, M., Hebert, P. D. N. & Mutanen, M. (2014) Barcoding beetles: a regional survey of 1872 species reveals high identification success and deep interspecific divergences. *PLoS One* **9**: e108651.

Piper, R. W. & Compton, S. G. (2003) Subpopulations of *Cryptocephalus* beetles (Coleoptera: Chrysomelidae): geographically close but genetically far. *Diversity and Distributions* **9**: 29-42.

Plewa, R., Sikora, K., Gutowski, J., Jaworski, T., Tarwacki, G., Tkaczyk, M., Rossa, R., Hilszczański, J., Magoga, G. & Kajtoch, Ł. (2018). Morphology, genetics and *Wolbachia* endosymbionts support distinctiveness of *Monochamus sartor sartor* and *M. s. urussovii* (Coleoptera: Cerambycidae). *Arthropod Systematics & Phylogeny* **72**: 123-135.

Popescu, A. A., Huber, K. T. & Paradis, E. (2012) ape 3.0: new tools for distance based phylogenetics and evolutionary analysis in R. *Bioinformatics* **28**: 1536-1537.

Potter, C., Tang, C. Q., Fonseca, V., Lallias, D., Gaspar, J. M., Thomas, K. Creer, S. & Potter, S. (2017). De novo species delimitation in metabarcoding datasets using ecology and phylogeny. *Peer J. Preprints* **5**: e3121v1.

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

Raupach, M. J., Hannig, K., Morinière, J. & Hendrich, L. (2016) A DNA barcode library for ground beetles (Insecta, Coleoptera, Carabidae) of Germany: The genus *Bembidion* Latreille, 1802 and allied taxa. *ZooKeys* **592**: 121-141.

Sassi, D. (2001) Nuove specie del genere *Cryptocephalus* vicine a *Cryptocephalus marginellus* (Coleoptera Chrysomelidae). *Memorie della Societa Entomologica Italiana* **80**: 107-138.

Sassi, D. (2011) A new species of the *Cryptocephalus marginellus* complex from Italian Western Alps (Coleoptera: Chrysomelidae: Cryptocephalinae). *Genus* **22**: 123-132.

Sassi, D. (2014) Taxonomic remarks, phylogeny and evolutionary notes on the leaf beetle species belonging to the *Cryptocephalus sericeus* complex (Coleoptera: Chrysomelidae: Cryptocephalinae). *Zootaxa* **3**: 333-378.

Sawadogo, A., Nagalo, E., Nacro, S., Rouamba, M. & Kenis, M. (2015) Population dynamics of *Aphthona whitfieldi* (Coleoptera: Chrysomelidae), pest of *Jatropha curcas*, and environmental factors favoring its abundance in Burkina Faso. *Journal of Insect Sci*ence **15**: 108.

Scherer, O. (1971) Das Genus *Livolia* Jacoby und seine umstrittene Stellung um System. *Entomologische Arbeiten aus dem Museum* **22**: 1-37.

Smith, M. A., Bertrand, C., Crosby, K., Eveleigh, E. S., Fernandez-Triana, J., Fisher, B. L., Gibbs, J., Mehrdad Hajibabaei, M., Hallwachs, W., Hind, K., Hrcek, J., Da-Wei Huang, D., Janda, M., Janzen, D. H., Li, Y., Miller, S. E., Packer, L., Quicke, D., Ratnasingham, S., Rodriguez, J., Rougerie, R., Shaw, M. R., Sheffield, C., Stahlhut, J. K., Steinke, D., Whitfield, J., Wood, M. & Zhou, X. (2012) *Wolbachia* and DNA Barcoding Insects: Patterns, Potential, and Problems. *PLoS One* **7**: e36514.

Szűcs, M., Schaffner, U., Price, W. J. & Schwarzländer, M. (2012) Post-introduction evolution in the biological control agent *Longitarsus jacobaeae* (Coleoptera: Chrysomelidae). *Evolutionary Applications* **5:** 858-868.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology Evolution* **30**: 2725-2729.

Thormann, B., Ahrens, D., Marín Armijos, D., Peters, M. K., Wagner, T. & Wägele, J. W. (2016) Exploring the Leaf Beetle Fauna (Coleoptera: Chrysomelidae) of an Ecuadorian Mountain Forest Using DNA Barcoding. *PLoS One* **11:** e0148268.

Warchalowski, A. (2003) Chrysomelidae. The Leaf-beetles of Europe and the Mediterranean area. *Natura Optima Dux Fundation.*

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**. e30490.

Warchalowski, A. (2010) The Palearctic Chrysomelidae: identification keys. Natura optima dux Foundation.

Wiemers, M. & Fiedler, K. (2007) Does the DNA barcoding gap exist? - a case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology* **4**: 8.

Wickham, H. (2009) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Winter, D. (2017) rentrez: Entrez in R. R package version 1.1.0. https://CRAN.R-project.org/package=rentrez.

2.1.4 Appendix

Suplementary information available at https://doi.org/10.1038/s41598-018-31545-9.

2.1.5 Personal contribution to the work

Samples collection in collaboration with M.M. Coinceving of the study in collaboration with M.M. and D.F. Bioinformatics analyses and writing of the manuscript in collaboration with M.M.

2.2 Research article

DNA Barcoding as useful tool to identify crop pest flea beetles of Turkey

Coral Sahin, D., Magoga, G., Ozdikmen, H. & Montagna, M. (2018) DNA Barcoding as useful tool to identify crop pest flea beetles of Turkey. *Journal of Applied Entomology* **143:** 105-117. doi: 10.1111/jen.12566

2.2.1 Summary

Flea beetles (Chrysomelidae: Galerucinae: Alticini), with ~8000 species worldwide, include pest species causing substantial economic damage to crops. The genera *Phyllotreta* and *Chaetochnema* include both pest and not-pest species. An accurate and fast taxonomic identification approach is required for discriminating among taxa for non-expert taxonomists; moreover, the utility of this approach spans from biodiversity conservation to the monitoring of pest species. DNA barcoding represents a reliable and easy identification tool based on the use of short DNA sequences. In this study, 45 new cox1 sequences of 13 *Phyllotreta* and five *Chaetocnema* species, representing ~30% and ~20% of the Turkish species belonging to these genera, were provided. These sequences increased by ~18% and ~25% the number of species of these genera whose sequences are available in BOLD. In order to test DNA barcoding efficiency in *Phyllotreta* and *Chaetocnema* species identification, we created a data set consisting of sequences belonging to species present in the Middle East and available in BOLD plus the sequences developed in this study (36 species). The efficiency of species identification, estimated using Best Closed Match analysis (with the *ad hoc* calculated optimal distance threshold of 1.5%), was 99%. The overall intraspecific and interspecific mean nucleotide divergences were 1.4% and 20%, respectively. Interestingly, cox1 sequences of *Phyllotreta nigripes* clustered into two well-separated groups with a high value of the between-group nucleotide distance (11.4%), which suggests the presence of cryptic species. In addition, information was provided on the crops exploited by the collected organisms and the observed damage.

2.2.2 Manuscript

Introduction

Flea beetles (Chrysomelidae: Galerucinae: Alticini), with approximately 500 genera and more than 8,000 species worldwide, are the largest tribe of Chrysomelidae (Coleoptera). In the Palearctic region they are represented by 90 genera and 1,388 species (Konstantinov & Vandenberg, 1996). To date in Turkey, 22 genera and 314 species have been recorded (Özdikmen et al., 2014; Bayram & Aslan, 2015; Aslan & Alkan, 2015). The species belonging to this group are small in size (usually of length 0.5-4.0 mm) and are characterized by enlarged hind femora due to the presence of the Maulik's organ, which confers the jumping capability (Scherer, 1971). Among the Alticini are listed species considered pests of crops worldwide (Metspalu et al., 2014). For example, among the species of the genus *Epitrix* Foundras, four are listed as quarantine species by European and Mediterranean Plant Protection Organization (EPPO), due to the damage caused to potatoes (Germain et al., 2013). As well as crop pests, the alien *Luperomorpha xanthodera* Fairmaire, 1888 is well know for the aesthetic damage caused to ornamental plants (Bene & Conti, 2009). In general, the economic losses are mainly caused by the overwintering adults emerging in spring and damaging the cotyledon and the first true leaves, which leads to the death of the plant (Bodnaryk & Lamb, 1991; Tansey et al., 2009; Metspalu et al., 2014). When the intensity of the flea beetle infestation is high, and the weather is favorable to their proliferation, the entire crop may even be lost (Knodel et al., 2004). The importance of flea beetles as pests, besides of being related to direct damage, is also related to their capability to transmit plant disease agents, such as viruses and bacteria (Booth et al., 1990). Insects retain the bacteria in the digestive organs during the winter diapause, and then in spring reinfect the plants while feeding on them. For example, *Chaetocnema denticulata* (Illiger, 1807) and *Chaetocnema pulicaria* Melsheimer, 1847 are vectors of *Pantoea stewartii* (Smith, 1898), the causative agent of Stewart's wilt, listed among the quarantine organisms by EPPO (Jourdheuil, 1963; Wensing et al., 2010). A further example is represented by the Rice yellow mottle virus, which is known to be transmitted by *Chaetocnema abyssinica* Jacoby, 1907, *Chaetocnema kenyensis* Bryant, 1948 and *Chaetocnema pulla* Chapuis, 1879 (Bakker, 1974). Some *Phyllotreta* Chevrolat species, especially *Phyllotreta striolata* (Fabricius, 1803) and *Phyllotreta cruciferae* (Goeze, 1777), were reported as vectors of the Turnip yellow mosaic virus, which poses a threat to cruciferous vegetables (Stobbs et al., 1998). On the other hand, due to their efficacy in damaging plants, some flea beetles belonging to the genera *Longitarsus* Berthold, *Altica* Müller, *Aphthona* Chevrolat and *Chaetocnema* Stephens were successfully adopted as biocontrol agents for weeds (Jonsen et al., 2001; Konstantinov, 1998; Konstantinov et al., 2001; Aslan & Gök, 2006; Cagáň et al., 2006). One exemplary application is represented by *Chaetocnema tibialis* (Illiger, 1807) which was adopted for controlling the invasive wild amaranth (Cagáň et al., 2006).

Note that even if a flea beetle can be considered harmful, its pest status may vary across the range of distribution in accordance with the varying intensity of the damage.

Chaetocnema, one of the two genera targeted by this study, has a worldwide distribution and includes species that are considered crop pests. *Chaetocnema concinna* (Marcsham, 1802) severely damages sugar beet, and for this reason, it is considered an important pest of this crop (EPPO, 2007). *C. tibialis* is regarded as a serious pest of this crop in Uzbekistan (Toreniyazov, 1999) and in Japan (Honma & Akiyama, 1981); *Chaetocnema basalis* is responsible for the 60% of the foliage infestation on seed potatoes in India (Thakur & Kashyap, 1994); *Chaetocnema aridula* (Gyllenhal, 1827) is a pest of barley (*Hordeum* spp.), oats (*Avena sativa*), rye (*Secale cereale*) and wheat (*Triticum* spp.); *Chaetocnema picipes* Stephens, 1831 of beet (*Beta vulgaris*) and *Chaetocnema hortensis* (Geoffroy, 1785) of barley *(Hordeum* spp.) flax (*Linumus itatissimum*) and wheat (*Triticum* spp.) (Aslan & Özbek, 1998; Konstantinov et al., 2011). The cosmopolitan genus *Phyllotreta*, includes species well known to be pests of cultivated Brassicaceae (Verdyck, 1998; Seeno & Wilcox, 1982). *P. cruciferae* and *P. striolata* cause, on average, the 10% of the annual yield loss of *Brassica rapa* L. and *Brassica napus* L. in North America (Lamb & Turnock, 1982; Verdyck, 1999). In Turkey, *Phyllotreta atra* (Fabricius, 1775), *Phyllotreta nigripes* (Fabricius, 1775), *Phyllotreta undulata* Kutschera, 1860 and *Phyllotreta vittula* (Redtenbacher, 1849) are important pests of cabbage (*Brassica oleracea*), radish (*Raphanus raphanistrum*), turnip (*Brassica rapa*) and rape (*Brassica napus*) (Aslan & Özbek, 2000). Some species of these two genera are considered pests causing severe damage to some crops, while on others they do not cause a significant economic damage. Among others, the case of *C. concinna* is worth noting. This is considered a serious pest on sugar and fodder beet but does not cause significant economic damage on potato, strawberry, buckwheat, spinach or sorrel, or in vineyards (Cagáň et al., 2000; Lesage & Majka, 2010). A further example is *P. vittula*, declared a pest only on maize, despite the wide host-plants spectrum that includes other cereals, Brassicaceae and Amaranthaceae (Szenási & Marcó, 2015; Cagáň et al., 2006). Moreover, some species belonging to *Chaetocnema* and *Phyllotreta* have a large spectrum of host plants including both cultivated and wild plants and can be accidentally found on crops only because they feed on the spontaneous vegetation growing in the surroundings or on inter-row weeds. Because of their similar external morphology and their small size, it is hard to distinguish pest from non-pest species directly in the field.

Flea beetle species present in crops are still managed using a broad spectrum of insecticides, regardless of whether an accurate taxonomic identification or an accurate evaluation of the incidence on crops has been made. The excessive use of chemical pesticides, besides threatening biological diversity, leads to environmental and health hazards; in addition, performing a high

number of unjustified treatments can lead the pest to develop insecticide resistances (Tangtrakulwanic et al., 2014). In order to adopt smart agricultural management practices, it is essential to be able to correctly and quickly identify the candidate pest. Within the integrated pest management approach, organism identification is the first essential step for allowing monitoring (abundance and the type of damage caused) and subsequently determining, if necessary, the economic threshold and the best management strategy (e.g., Radcliffe et al., 2009). Morphological identification requires time and high taxonomical expertise, and usually only allows the identification of organisms at the adult stage. Molecular methods, such as DNA barcoding (Hebert et al. 2003; Cywinska et al., 2003), offer the possibility of overcoming these limitations and allow organism identification at preimaginal and adult stages using a short DNA sequence as a barcode (Zhang et al., 2016). The approach has been successfully adopted in a plethora of studies spanning a wide range of target organisms, from vertebrates, (e.g., Schäffer et al., 2017; Britz et al., 2017) to plants, (e.g., Chattopadhyay et al., 2017) and insects, (e.g., Massimino Cocuzza & Cavalieri, 2014; Šigut et al., 2017; Khan et al., 2017; Wang et al., 2017; Montagna et al., 2016) and it has also been used in pests biosurveillance (Jalali et al., 2015; Ashfaq & Hebert, 2016).

The aims of this study are: *i*) to develop a reference database of cox1 sequences for the identification of *Chaetocnema* and *Phyllotreta* species present in Turkey; *ii*) to test the DNA barcoding approach efficiency for the identification of the species of the two genera. The developed database, useful for discriminating among pests and non-not pests *Chaetocnema* and *Phyllotreta* species in Turkey and neighbouring countries, will be helpful as decision support tool for monitoring the species and establishing effective and eco-compatible pest management strategies. In addition, a table of the host plants of the species included in this study and the information about the damage they caused, is provided.

Materials and Methods

Sampling, specimen manipulation and morphological identification

The *Phyllotreta* and *Chaetocnema* specimens examined in this study were collected from various natural and agricultural areas of Turkey (Ankara, Amasra, Bartın, Kayseri, Şanlıurfa and Zonguldak provinces) between 2014 and 2017 using a sweep net and aspirator (Table 1; Fig. 1). Some of the collected specimens were preserved in absolute ethanol and stored at -20°C soon after their collection, while others were pinned and dry preserved. The specimens were identified to the genus level by stereomicroscope (Wild M3, Wild Heerbrugg). After the DNA extraction, the specimens were dissected to remove the genitalia, then mounted on pins together with the genitalia and

identified at the species level using identification keys (Doguet, 1994; Warcholowski, 2010) and the collection of Nazife Tuatay Plant Protection Museum (8,000 specimens, 500 species of Chrysomelidae) as references. All the identifications were then confirmed by Carlo Leonardi (Museo Civico di Storia Naturale di Milano, Italy).

During the collecting campaigns, information was recorded on the host plants exploited and the damage caused by the insect to crops. We defined three categories of damage based on the qualitative status of the plant leaves observed in the field: i) no damage, the plants on which the insects were collected had no damage on leaves; ii) low damage, the plants on which the insects were collected presented few holes on the leaves, usually concentrated in a limited area (Supplementary Fig. 1 A); and iii) high-damage, the plants on which the insects were collected presented holes and scrapes widespread on the leaf surfaces, presence of necrotic areas, wilting of the leaves (Supplementary Fig. 1 B).

DNA extraction and PCRs

DNA was extracted using the phenol-chloroform method (Doyle & Doyle, 1990), with the slight modifications reported by Mereghetti et al., 2017, starting from the whole body, removing the abdomen and crushing three legs (a foreleg, mid leg and hind legs). The extracted DNA was quantified using a NanoDrop ND-1000 spectrophotometer, and 20-40 ng sample were used as template for the PCR amplification. The classic barcoding region of the mitochondrial cox1 gene (658 bp) was amplified by PCR using the universal primers for metazoans LCO1490 and HCO2198 (Folmer, 1994). PCR reactions were performed in a final volume of 25 µl using the concentration of reagents and PCR thermal profile used by Montagna et al., 2017. Successful amplifications were determined by gel electrophoresis and PCR products were sequenced by ABI technology (Applied Biosystems, Foster City, CA, USA). The obtained electropherograms were manually edited and checked using Geneious Pro 5.1 (Biomatters Ltd., Auckland, New Zealand).

DNA barcoding analyses

The obtained cox1 gene sequences were aligned using MUSCLE (Edgar, 2004) implemented in MEGA 6 (Tamura et al., 2013), which allows the maintenance of the codons structure. Two datasets were built: *ds1* composed of cox1 sequences developed in this study and *ds2* composed of cox1 sequences of *Phyllotreta* and *Chaetocnema* species listed in the fauna of Turkey and of neighbouring countries mined from BOLD (http://www.boldsystems.org/index.php) plus the sequences of *ds1*.

Intraspecific and interspecific mean distances were calculated with the R library "spider" (Brown et al., 2012) starting from a pairwise nucleotide distance matrix based on Kimura's two-parameter (K2P) substitution model (Kimura, 1980), considered a suitable model when p-distances between sequences are low (Nei & Kumar, 2000). All the *cox1* sequences obtained in this study were clustered using the Barcode Index Number (BIN) in BOLD (Ratnasingham & Hebert, 2013). Due to the absence in BOLD of many of the species for which we produced *cox1* sequences, we made a further test of the reliability of the data set through a Best Closed Match analysis (Meier et al., 2006) performed on *ds2* using the R package "spider" (Brown et al., 2012). This method compares each sequence within a data set with the others, in order to test if the best matches (i.e., the closet pairs of sequences in term of nucleotide distance), occur between sequences of organisms belonging to the same morphological species. Furthermore, the nucleotide distance threshold that minimizes the cumulative error was calculated using the Optimal Threshold analysis in the "*spider*" package (Brown et al., 2012). This analysis tests different values of the nucleotide distance threshold and indicates the one for which the sum of false positives (FP, conspecifics with a value of nucleotide divergence higher than the threshold value) plus false negatives (FN, heterospecifics with a value of nucleotide divergence lower than the threshold value), i.e., the cumulative error, is lowest. The obtained threshold value was also used in the Best Closed Match analysis, which requires the highest limit value for the intraspecific distance to be declared. In order to confirm the results obtained from the Best Close Match and BIN analyses, a distance-based neighbor-joining tree was inferred using SeaView version 4 (Gouy et al., 2010) on ds1 with the addition of one orthologous sequence for each of the species included in this data set and available in BOLD. *Altica oleracea* (Linnaeus, 1758) was included as the outgroup. Kimura's two-parameters (1980) was used as the model of nucleotide substitution and 100 bootstrap replicates were performed. Finally, the sequences were deposited in the BOLD system (accession numbers from MH407425 to MH407469).

Table 1. Flea beetle specimens analysed in the study.

Species	Collecting Locality	Specimen[a]	Altitude[b]	Lat N[c]	Long E[d]	Date[e]
P. cruciferae	Ankara, Haymana, Soğulca	1	948	39°22'26"	32°21'03"	07.v.2015
	Ankara, Ayaş, Feruz	1	1075	40°14'21"	32°29'18"	13.viii.2015
	Kayseri, Kocasinan, Boğazköprü	1	1040	38°46'01"	35°17'01"	23.vi.2016
	Kayseri, Bünyan, Malatya way	3	1407	38°49'03"	35°55'17"	21.ix.2016
	Ankara, Beypazarı, Kayabükü	1	512	40°04'77"	31°48'77"	17.v.2017
P. nigripes	Ankara, Beypazarı, Köşebükü	3	493	40°04'47"	31°48'44"	16.iv.2014
	Ankara, Beypazarı, Kayabükü	1	512	40°04'47"	31°48'44"	17.v.2017
	Kayseri, Kocasinan, Himmetdede	1	1211	38°55'23"	35°04'42"	19.ix.2016
	Kayseri, Kocasinan, Himmetdede	2	1250	38°54'36"	35°05'18"	19.ix.2016

	Ankara, Ayaş, Sinanli	2	808	40°00'52"	32°17'47"	17.v.2017
P. atra	Bartın, Güzelcehisar	2	237	41°38'54"	32°12'30"	12.v.2015
	Zonguldak, Devrek, Bakırcılar	1	124	41°15'60"	31°58'98"	02.vi.2016
	Zonguldak, Devrek, Bakırcılar	1	124	41°15'60"	31°58'98"	12.iv.2017
P. procera	Kayseri, Kocasinan, Himmetdede	1	1215	38°55'23"	35°04'42"	19.ix.2016
	Kayseri, Kocasinan, Himmetdede	1	1250	38°54'36"	35°05'18"	19.ix.2016
	Kayseri, Özvatan	1	1337	39°06'47"	35°42'50"	21.ix.2016
P. astrachanica	Zonguldak, Beycuma, Bölücek	1	274	41°19'31"	31°57'85"	12.iv.2017
P. diademata	Bartın, Karasu	1	36	41°39'21"	32°14'07"	28.iv.2017
P. striolata	Zonguldak, Devrek, Bakırcılar	1	124	41°15'60"	31°58'98"	15.vi.2016
	Zonguldak, Devrek, Bakırcılar	1	124	41°15'60"	31°58'98"	12.iv.2017
P. nemorum	Bartın, Karasu, Center	1	37	41°39'23"	32°14'01"	28.iv.2017
	Zonguldak, Devrek, Bakırcılar	1	124	41°15'60"	31°58'98"	12.iv.2017
P. corrugata	Şanlıurfa, Akçakale, T.Demirören	1	382	36°54'04"	38°55'03"	30.iv.2014
P. undulata	Bartın, Ulus, İğneciler	1	548	41°38'50"	32°35'94"	12.iv.2017
P. pallidipennis	Kayseri, İncesu, Örenşehir	1	1034	38°42'21"	38°15'27"	20.vi.2016
P. varipennis	Bartın, Güzelcehisar	1	237	41°38'54"	31°12'30"	12.v.2015
	Zonguldak, Devrek	1	153	41°13'05"	31°56'53"	26.v.2015
P. erysimi	Ankara, Beypazarı, Kayabükü	1	512	40°04'77"	31°48'77"	17.v.2017
C. tibialis	Ankara, Beypazarı, Dibecik	1	534	40°06'07"	32°04'25"	29.v.2014
	Kayseri, Kocasinan, Himmetdede	1	1250	39°06'47"	35°42'50"	19.ix.2016
	Kayseri, Özvatan	1	1337	38°54'36"	35°05'18"	21.ix.2016
	Zonguldak, Devrek, Bakırcılar	1	103	41°15'60"	31°58'98"	12.iv.2017
	Bartın, Kozcağız, Merkez	1	103	41°29'28"	32°20'19"	11.v.2017
C. concinna	Amasra, Makaracılar	1	187	41°43'68"	32°26'66"	11.v.2017
C. coyei	Kayseri, Bünyan, Malatya way	2	1407	38°49'03"	35°55'17"	21.ix.2016
C. arenacea	Kayseri, Bünyan, Malatya way	1	1407	38°49'03"	35°55'17"	21.ix.2016
C. conducta	Amasra, Makaracılar	1	187	41°43'68"	32°26'66"	11.v.2017

[a]Number of collected specimens; [b]altitude of the collecting point; [c]latitude of the collecting point; [d]longitude of the collecting point; [e]date of collection.

Results

Host plant information and crop damage

A total of 45 specimens attributed to 18 *Phyllotreta* and *Chaetocnema* species were collected from 37 Turkish localities (Table 1; Fig. 1). *C. tibialis* was found in high numbers on chard (*Beta vulgaris cicla*), sugar beet *(Beta vulgaris saccharifera)*, beet *(Beta vulgaris)*, radish (*Raphanus sativus)* and cress (*Lepidium sativum*) (Table 2, Supplementary Table 1). The recorded high level of damage, consisting of a high number of holes spread over the whole leaf surface (Fig. 2 A), suggests that this species could be considered a pest. Despite the high number of individuals on spinach (*Spinacia oleracea*), bean (*Phaseolus vulgaris*) and cabbage (*Brassica* sp.), significant damage was not observed. A few individuals were also found on parsley (*Petroselinum crispum*),

tomato (*Solanum lycopersicum*), pepper (*Capsicum* sp.), lettuce (*Lactuca sativa*) and courgette (*Cucurbita* sp.), but were not observed feeding on them.

Regarding *Phyllotreta*, seven species (*P. atra, P. cruciferae, P. nigripes, P. undulata*, *Phyllotreta erysimi* Weise, 1900, *Phyllotreta varipennis* (Boieldieu, 1859), *P. striolata*) were collected on crops. Cress, cabbage, radish, kale (*Brassica oleracea acephala*), broccoli (*Brassica oleracea italica*), rocket (*Eruca vesicaria*) and cauliflower (*Brassica oleracea botrytis*) showed a high level of damage (Fig. 2 B); a large number of individuals were observed feeding. In addition, some individuals were observed on pepper, spinach, lettuce and potato, but no significant damage was recorded (Table 2, Supplementary Table 1).

Datasets features and nucleotide distances

Data set *ds1* includes a total of 45 *cox1* sequences (size ranging from 481 to 642 bp) for 13 species of *Phyllotreta* (35 sequences) and five species of *Chaetocnema* (10 sequences) (Table 1; Supplementary Table 2), representing approximately 30% and 20% of the species known in Turkey (44 *Phyllotreta* and 26 *Chaetocnema* species). In data set *ds1*, a total of 8 and 25 haplotypes were recovered for *Chaetocnema* and *Phyllotreta*, respectively. For seven of these species, (i.e., *Phyllotreta pallidipennis* Reitter, 1891, *Phyllotreta corrugata* Reiche, 1858, *Phyllotreta diademata* Foudras, 1860, *P. erysimi, Chaetocnema coyei* (Allard, 1863), *Chetocnema arenacea* (Allard, 1860), *Chaetocnema conducta* (Motschulsky, 1838), no *cox1* sequences were previously available in either the BOLD or the GenBank databases (Fig. 3). In total, nine of the 18 species have unique BINs, and in the case of *P. nigripes* two BINs are present. Specimens identified as *P. cruciferae* clustered into a BIN that includes specimens assigned to *P. cruciferae* and *Phyllotreta albionica* (LeConte, 1857).

Data set *ds2* includes the 45 *cox1* sequences obtained in this study plus the sequences of *Phyllotreta* and *Chaetocnema* species present in Turkey and neighbouring countries available in BOLD (720 sequences, 139 haplotypes, 29 species), for a total of 765 *cox1* sequences and 36 species (Supplementary Table 3).

The analysis of pairwise K2P nucleotide distances in *ds1* showed mean intraspecific and interspecific distance values of 2.9% (range 0-12%) and 19% (range 8-27%) respectively (Supplementary Fig. 2), while the same metrics for *ds2* are of 1.4% (range 0-12%) and 20% (range 7.4-31%) respectively (Fig. 4 A). Pairwise nucleotide distances calculated only for the genus *Chaetocnema* (15 species, 162 sequences) showed mean intraspecific and interspecific distances of 0.15% and 16.5% respectively (Fig. 4 B); the same metrics were of 1.5% and 18% in the case of the genus *Phyllotreta* (21 species, 603 sequences; Fig. 4 C). The distribution of K2P pairwise distances

highlights the presence of a clear barcode gap within and between species (Fig. 4 A, Supplementary Fig. 2). One exception is represented by *P. nigripes* (33 sequences), where intraspecific pairwise comparisons reached nucleotide distance values of ~12% (Supplementary Fig. 3). The pairwise nucleotide K2P distances were calculated also for the *P. nigripes* sequences present in *ds2*. Two separate groups of sequences were identifiable, with a mean intragroup K2P nucleotide distances of 0.38% (range 0-1.23%; *n* = 27) and of 0.45% (range 0-0.1%; *n* = 6). The mean K2P nucleotide distance between the two groups is of 11.4% (range 10.9-12.1%; *n* = 33). BIN assignments of sequences from Turkey confirmed the existence of two *P. nigripes* BINs (ACR44611 and ACA7757). Specimens assigned to ACA7757 were collected in Kirmir valley (west side of the Central Anatolian Region). Specimens assigned to ACR44611 were collected in a steppe vegetation area on the east side of the Central Anatolian Region.

Figure 1. Organisms collecting sites. Red dots indicate each specimen collecting site, color intensity is proportional to number of specimens per site.

Optimal threshold and best close match analyses

Optimal threshold estimation for *ds2* resulted in a nucleotide distance value of 1.5%, with an associated cumulative error of one (FN = 0; FP = 1). The false positive was associated with a sequence of *C. aridula* mined from the BOLD database. In BOLD, there are nine public sequences assigned to *C. aridula*, associated with two BIN numbers (ABZ9838 and ACC5699). In Best Closed Match analysis, one of these sequences of *C. aridula* did not match any conspecific sequences. The pairwise K2P nucleotide distances analysis showed that this sequence has a distance of 12.2% from those of conspecifics.

The efficiency of the DNA barcoding calculated for *ds2*, using 1.5% for the optimal threshold, was ~99%. The 1% error in the analysis is not related to the sequences developed in this study; out of the 18 species included in this study, 10 had a conspecific as best close match, showing complete concordance with morphological identifications. Eight species were represented by only one DNA barcode and, as may be expected, no match with other congeneric species was obtained below the threshold. The neighbor-joining tree confirmed the results previously obtained on the basis of the Best Close Match and the BIN analyses, and all the species possessing at least one conspecific were monophyletic and supported with high bootstrap values (100; Fig. 5). Interestingly, even if *P. striolata* was monophyletic, the two organisms collected in Turkey were connected by a long branch to the organism collected in Canada, for which the *cox1* sequence is available on BOLD (mean pairwise nucleotide divergence of 4.7%). Concerning *P. nigripes*, two well-supported clades were recovered (bootstrap = 100; Fig. 5), confirming the results of the BIN analysis.

Table 2. Host plants information and the level of damage for the species collected during 2014-2016 collecting campaigns on cultivated fields.

Species	Host plant species	Damage[a]	Reference
C. tibialis	*Beta vulgaris cicla, B. vulgaris saccharifera, B.vulgaris, Raphanus sativu, Lepidium sativum*	HD	Field observation*
	Spinacia oleracea, Lepidium sativum, Brassica sp.	LD	Field observation*
	Beta vulgaris saccharifera	HD	Yildirim & Özbek, 1992
P. atra	*Lepidium sativum, Brassica* sp., *Raphanus sativus, Brassica oleracea acephala, Eruca vesicaria*	HD	Field observation*
	Raphanus sativus, Brassica rapa	HD	Aslan & Özbek, 1998
P. cruciferae	*Lepidium sativum, Brassica* sp., R*aphanus sativus, B. oleracea acephala, Eruca vesicaria*	HD	Field observation*
	Brassica sp, B. oleracea acephala, B. oleracea botrytis, B. oleracea italica	HD	Sáringer, 1990; Vig, 1992; Balázs et al., 1998; Spilák et al., 1998; Vörös & Garamvölgyi, 1998
	Raphanus sativus, Brassica rapa	HD	Aslan and Özbek, 1998;
	Brassica napus, Brassica rapa, Brassica juncea	HD	Burgess, 1982; Soroka et al., 2005; Ulmer & Dosdall, 2006; Dosdall & Mason 2010; Metspalu et al., 2014
P. erysimi	*Raphanus sativus*	HD	Field observation*
	Brassica sp.	LD	Aslan and Özbek, 1998
	Raphanus sativus, Brassica rapa	HD	
P. nigripes	*Lepidium sativum, Brassica* sp., *Raphanus sativus, Brassica oleracea italic, Brassica oleraceaa cephala, Eruca vesicari*	HD	Field observation*

	Brassica sp., *Raphanus sativus*, *Brassica rapa*	HD	Kaszab, 1962; Sáringer, 1990; Aslan & Özbek, 1998
P. striolata	*Brassica* sp.	HD	Field observation*
P. undulata	*Brassica* sp.	HD	Field observation*
	Solanum tuberosum L.	ND	
	Brassica sp.	HD	Aslan & Özbek, 1998; Brelih et al., 2003; Metspalu et al., 2014
	Brassica napus oleifera, *Brassica rapa oleifera*	HD	Ekbom, 1990; Alford et al., 2003; Veromann et al., 2006
P. variipennis	*Brassica* sp., *Raphanus sativus*	HD	Field observation*

[a]HD: high damages; LD: low damages; ND: no damages but individuals present with high number on the plant. *The field observations were performed within the project "Determination of Host Plants of Alticini Species on Vegetables".

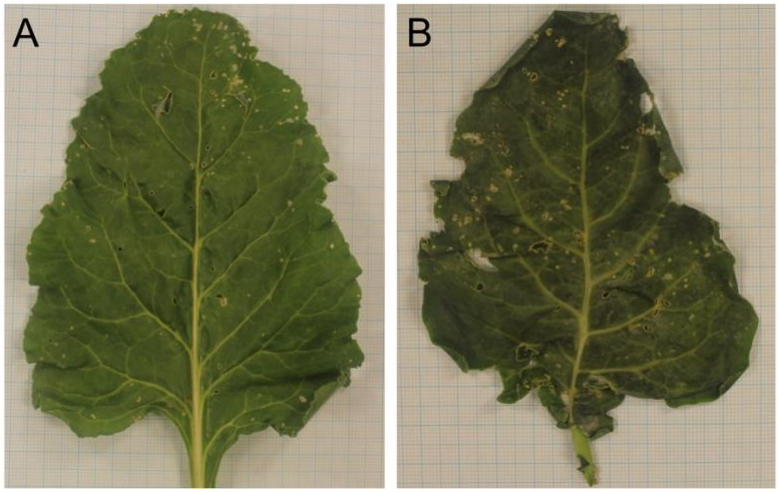

Figure 2. Images of high damage caused by *Chaetocnema* and *Phyllotreta* species to crops. A: damage caused by *Chaetocnema tibialis* to a sugar beet leaf. B: damage caused by *Phyllotreta cruciferae* to a cauliflower leaf.

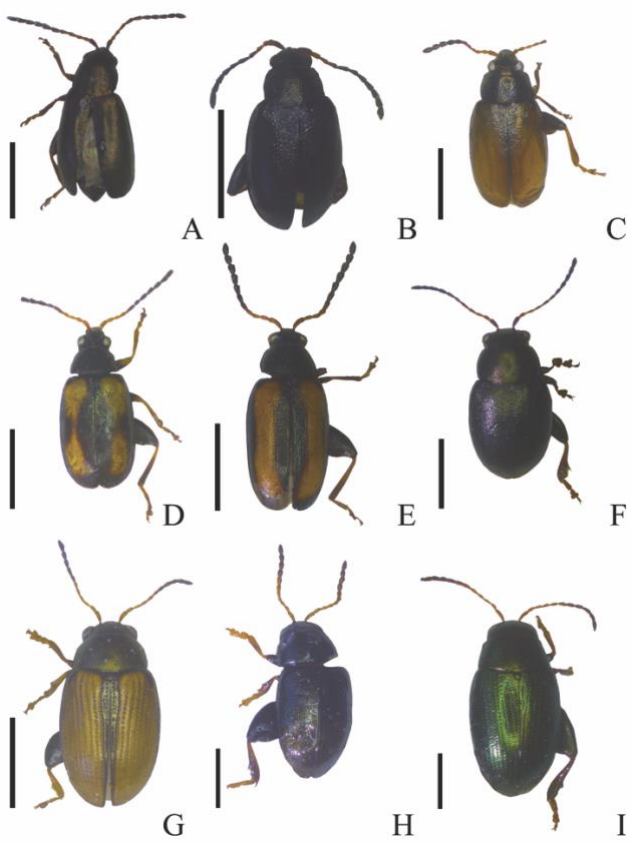

Figure 3. Images of specimens processed in this study for which *cox1* sequences were not already available on BOLD. A: *Phyllotreta corrugata*; B: *P. astrachanica*; C: *P. pallidipennis*; D: *P. varipennis*; E: *P. erysimi*; F: *Chaetocnema aranecae*; G: *C. conducta*; H and I: *C. coyei*. Scale bars size are 1 mm.

Discussion

In this study, 45 *cox1* barcodes of two flea beetles genera that include pest species present in the Middle East region, namely *Phyllotreta* and *Chaetocnema*, were obtained. In addition, the effectiveness of the developed data set as a tool for a fast and accurate taxonomic identification was tested. Information on the host plants and on the pest status of some of the collected species was also provided to improve the knowledge of the species biology and allow the identification of those economically important species.

With regard to host plants, *C. tibialis* is the only species of the genus included in this study that was found feeding on crops; it was observed to be one of the most abundant and common species in the field, resulting in serious damage on chard (*Beta vulgaris cicla*), sugar beet (*Beta vulgaris saccharifera*), beet (*Beta vulgaris*), radish (*Raphanus raphanistrum*) and cress (*Lepidium sativum*). With the exception of sugar beet, for which *C. tibialis* has been previously reported to be a destructive pest in Europe and in Turkey (Pataki, 1967; Stef & Buzinovschi, 1982; Neves, 1983;

Yıldırım & Özbek, 1992; Aslan & Özbek, 1998; Cagáň et al., 2000; Cagáň et al., 2006), the remaining crops are reported here for the first time as host plants in cases where the species reaches the pest status.

Within the genus *Phyllotreta*, *P. atra*, *P. cruciferae* and *P. nigripes* were collected in large numbers on Brassicaceae crops, where significant damage was observed. Some individuals were also found on bean (*Phaseolus vulgaris*), lettuce (*Lactuca sativa*), pepper (*Capsicum* sp.), spinach (*Spinacia oleracea*) and courgette (*Cucurbita* sp.), but no damages was noticed. *Phyllotreta nigripes* and *P. cruciferae* have been known to be significant pests of crucifers, especially of cabbage (Kaszab, 1962; Sáringer, 1990) which supports our observation. *Phyllotreta undulata*, known to be one of the most destructive species of the genus and the most common flea beetle feeding on cabbage in several European countries (Brelih et al., 2003; Metspalu, et al., 2014), was found in a high numbers on cabbage, with significant damage. A comparable number of *P. undulata* were collected on potato close to cabbage fields, but no damage was observed. The collection of *P. variipennis* on cabbage represents the first record of this species as pest. *P. erysimi*, was observed causing high damages on radish, as reported also in previous studies (Aslan & Özbek, 1998). *Phyllotreta striolata*, known as a serious pest of cabbage, was found to be highly destructive in cabbage production areas. Despite its importance in agricultural areas, the knowledge concerning the host plants of this species is still fragmentary.

Our study confirmed the effectiveness of DNA barcoding as a molecular identification tool for the analysed genera. The identification efficiency through DNA barcoding showed a 99% accuracy for *ds2* (our data set plus available sequences of *Phyllotreta* and *Chaetocnema* species present in Turkey and neighbouring countries in the BOLD database (n = 765)). High correspondence between morphospecies and molecular species based on DNA barcodes was achieved, and the nucleotide distances within and between species highlighted the presence of an almost clear barcode gap. The intraspecific mean nucleotide distance value estimated in *ds1* is in agreement with those obtained in previous studies on Chrysomelidae (Montagna et al., 2013; Magoga et al., 2016; Kolasa et al., 2017). The values of the mean nucleotide intraspecific distance calculated for *Chaetocnema* and *Phyllotreta* genera separately are different, viz. 0.15% and 1.5%. Interestingly, the low value observed for *Chaetocnema* might be due to a high number of identical haplotypes, although the individuals were collected in different localities. Inconsistencies in the barcode gap are associated with *P. nigripes*, whose sequences are separated into two BINS (ACR44611 and ACA7757). The nucleotide distance between the members of these two groups is large (11.4%) and is within the range of interspecific nucleotide distances. In addition, the neighbor-joining tree confirmed the presence of two separated lineages of *P. nigripes*. The external morphology of the individuals

assigned to the two groups is identical, but it is interesting to note that the morphology of the spermatheca differs slightly. Moreover, the distance between the collection localities (~340 km) the differences in their host plants and in their topographic and habitat characteristics, could support the possible presence of a cryptic species. Further analyses are currently in progress to evaluate this possibility.

With respect to the clustering in a unique BIN number of *P. cruciferae* and *P. albionica* sequences in the BOLD database, *P. cruciferae*, is hardly distinguishable from *P. albionica*. The slight difference is that *P. cruciferae* has a blue lustre while *P. albionica* has a bronze lustre (Capinera, 2001). According to morphological identification and geographical data (*P. albionica* is limited to the Nearctic region), our sequences can be clearly identified as *P. cruciferae* and all the sequences assigned to that BIN probably belong to *P. cruciferae* also.

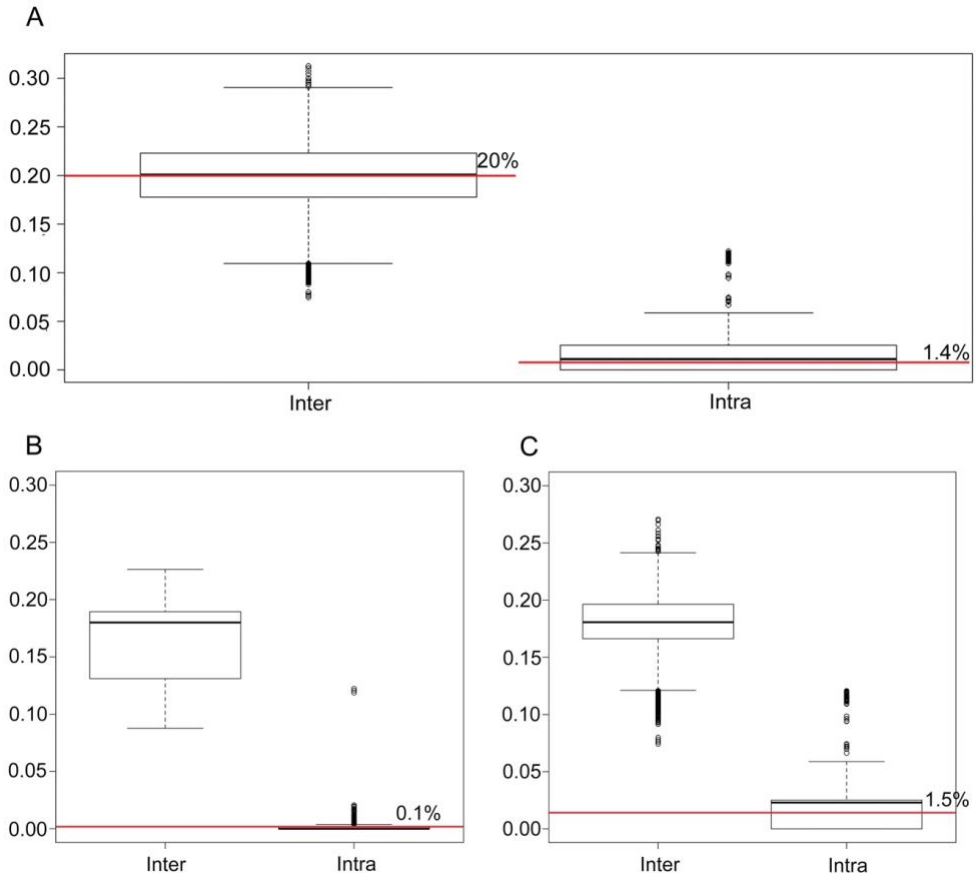

Figure 4. Boxplot representing intra-interspecific distances distribution. A: ds2; B: *Chaetocnema*; C: *Phyllotreta*. Red lines indicate mean values.

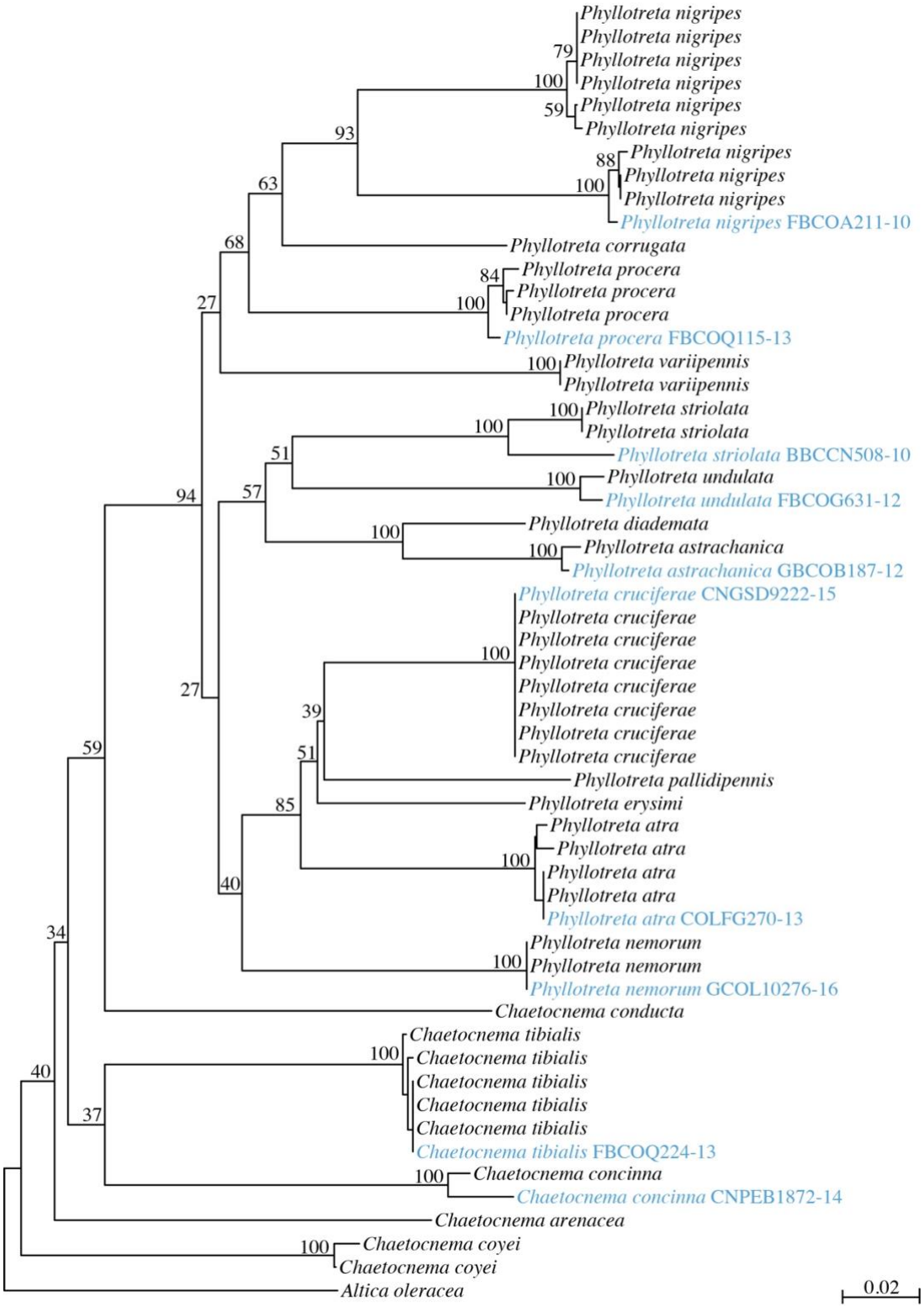

Figure 5. Neighbour-joining tree inferred on ds1 plus one representative sequence for each species available in BOLD. Bootstrap values are reported above the main lineages; the scale bar indicates the distance expressed in nucleotide substitutions per site; in light blue are reported the sequences from BOLD.

Conclusion

In this study, as well as providing information on the host plants and the pest status of the collected species, a *cox1* barcoding data set was developed and tested for the molecular identification of two flea beetle genera (*Phyllotreta* and *Chaetocnema*), which include pest species. Alticini includes species that are hardly distinguishable morphologically, and it is quite common that, among highly similar species, only some are crop pests and require the use of chemical pesticides for their management. Therefore, an accurate and relatively rapid approach for the taxonomic identification of individuals must to be regarded as pivotal for the development of economically and sustainable management strategies for crop pests, as well as for far-sighted policies regarding biodiversity conservation. Considering the promising results achieved in this study (99% of efficiency in species identification using *cox1* sequences), which represents the first step in the development of a DNA barcoding database for the leaf beetles of Turkey and neighbouring countries, we believe that a further effort in this field of research is vital.

2.2.3 References

Alford, D. V., Nilsson, C. & Ulber, B. (2003) Insect pests of oilseed rape crops. In Alford D. A. (Ed.) *Biocontrol of oilseed rape pests* (pp. 10-41). Oxford: Blackwell Science.

Ashfaq, M. & Hebert, P. D. (2016) DNA barcodes for bio-surveillance: regulated and economically important arthropod plant pests. *Genome* **59**: 933-945.

Aslan, E. G. & Alkan, K. (2015) The Alticini (Coleoptera: Chrysomelidae: Galerucinae) fauna of Davraz Mountain (Isparta): comments on host plant and altitude preferences with two new records for Turkish fauna. *Turkish Journal of Zoology* **9**: 488-493.

Aslan, E. G. & Gök, A. (2006) Host-plant relationships of 65 flea beetles species from Turkey, with new associations (Coleoptera: Chrysomelidae: Alticinae). *Entomological News* **117**: 297-308.

Aslan, İ. & Özbek, H., (1998) Erzurum' da Sebzelerde Zarar Yapan Yaprak Böcekleri (Coleoptera, Chrysomelidae). *Proceedings of the Third Vegetables Symposium*, 28-30 September 1998, Tokat, 335-340.

Aslan, İ. & Özbek, H. (2000) *Phyllotreta* (Coleoptera, Chrysomelidae, Alticinae) species, their biology and damage in some vegetables in Cruciferae family. *Proceedings of the Third Vegetables Symposium*, 11-13 September 2000, Isparta: 203-207.

Bakker, W. (1974) Characterization and ecological aspects of rice yellow mottle virus in Kenya. Wageningen. (Doctoral thesis).

Balàzs, K., Nagy, G. & Garamvölgyi, I. (1998) Plant protection of field-grown cabbages (common white cabbage, savoy, red cabbage). *Növényvédelem* **34**: 79-86.

Bayram, F. & Aslan, E. G., (2015) Comparation of Alticini (Coleoptera: Chrysomelidae: Galerucinae) species diversity in different habitats selected from Bafa Lake Natural. *Turkish Journal of Entomology* **39**: 147-157.

Bene, G. & Conti, B. (2009) Notes on the biology and ethology of *Luperomorpha xanthodera*, a flea beetle recently introduced into Europe. *Bulletin of Insectology* **62**: 61-68.

Bodnaryk, R. P. & Lamb, R. J. (1991) Mechanism of resistance to the flea beetle, *Phyllotreta cruciferae* (Goeze), in mustard seedlings, *Sinapsis alba* L. *Canadian Journal of Plant Science* **71**: 13-20.

Booth, R. G., Cox, M. L. & Madge, R. B. (1990) *Guides to Insects of Importance to Man, 3. Coleoptera IIE*. Cambridge, UK: CABI Publisher.

Brelih, S., Döberl, M., Drovenik, B. & Pirnat, A. (2003) Gradivo za favno hroščev (Coleoptera) Slovenije. 1. Prispevek. Polyphaga: Chrysomeloidea (Phytophaga): Chrysomelidae: Alticinae. Materialien zur

Käferfauna (Coleoptera) Slowenien. 1. Beitrag: Polyphaga: Chrysomeloidea (Phytophaga): Chrysomelidae: Alticinae. *Scopolia*. **50**: 1279.

Britz, R., Adamson, E., Raghavan, R., Ali, A. & Dahanukar, N. (2017) *Channa pseudomarulius***,** a valid species of snakehead from the Western Ghats region of peninsular India (Teleostei: Channidae), with comments on *Ophicephalus grandinosus***,** *O. theophrasti* and *O. leucopunctatus. Zootaxa* **4299**: 529-545.

Brown, A. 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**: 562-565.

Burgess, L. (1982) Occurrence of some flea beetle pests of parkland rapeseed crops in open prairie and forest in Saskatchewan (Coleoptera: Chrysomelidae). *Canadian Entomologist* **114**: 623-627.

Cagán, L., Vráblová, M. & Tóth, P. (2000) Flea Beetles (Chrysomelidae: Alticinae) Species Occurring on *Amaranthus* Spp. In Slovakia. *Journal of Central European Agriculture* **1**: 14-25.

Cagáň, L., Tóth, P. & Tóthová, M. (2006) Population dynamics of *Chaetocnema tibialis* Illiger and *Phyllotreta vittula* (Redtenbacher) on the weed *Amaranthus retroflexus* L. and cultivated *Amaranthus caudatus* L. *Plant Protection Science* **42**: 73-80.

Capinera, J. L. (2001) Handbook of Vegetable Pests. New York: Academic Press.

Chattopadhyay, P., Banerjee, G. & Banerjee, N. (2017) Distinguishing Orchid Species by DNA Barcoding: Increasing the Resolution of Population Studies in Plant Biology. *OMICS*: *A Journal of Integrative Biology* **21**: 711-720.

Doguet, S. (1994) Coléoptères Chrysomelidae Volume 2 Alticinae. *Faune de France, France et Regions Limitrophes*. Paris: Fédération Française des Sociétés de Sciences Naturelles.

Dosdall, L. M. & Mason, P. G. (2010) Key pests and parasitoids of oilseed rape or canola in North America and the importance of parasitoids in integrated management. Williams I. H. (Ed.), *Biocontrol-based Integrated Management of Oil seed Rape Pests* (pp. 167-213). Dordrecht: Springer.

Doyle, J. J. & Doyle, J. L. (1990) Isolation of plant DNA from fresh tissue. *Focus* **12:** 13-15.

Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.

EPPO [European and Mediterranean Plant Protection Organization] (2007). *Atomaria linearis, Thrips angusticeps, Chaetocnema tibialis, Chaetocnema concinna*. *OEPP/EPPO Bulletin* **37**: 33-35.

Ekbom, B. (1990) Flea beetles (*Phyllotreta* spp.) in spring oilseed rape in Sweden. *IOBC-WPRS Bull* **12**: 57-61.

Farrugia, C. (1997) Insect pests on cauliflower *(Brassca oleracea* var. *botrytls)* in Gozo (Maltese Islands, Central Mediterranean). *The Central Mediterranean Naturalist* **2**: 52-165.

Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3:** 294-299.

Germain, J. F., Chatot, C., Meusnier, I., Artige, E., Rasplus, J. Y. & Cruaud, A. (2013) Molecular identification of *Epitrix* potato flea beetles (Coleoptera: Chrysomelidae) in Europe and North America. B*ulletin of Entomological Research* **103**: 354-362.

Gouy, M., Guindon, S. & Gascuel, O. (2010) SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* **27**: 221-224.

Hebert, P. D. N., Cywinska, A., Ball, S. L. & deWaard, J. R. (2003) Biological identifications through DNA barcodes. *Royal Society* **270**: 313-321.

Honma, K. & Akiyama, Y. (1981) Observations on *Chaetocnema concinna* Marshall injurious to sugar beet and its related species *C. discreta* Baly (Coleoptera: Chrysomelidae). *Japanese Journal of Applied Entomology and Zoology* **25**: 123-125.

Jalali, S. K., Rakshit, O. & Venkatesan, T. (2015) DNA barcoding for identification of agriculturally important insects. Chakravaryhy, A. K. (Ed.), *New horizons in insect science: towards sustainable pest management* (pp. 13-22). New Dehli: Springer.

Jonsen, I. D., Bourchier, R. S. & Roland, J. (2001) The influence of matrix habitat on *Aphthona* flea beetle immigration to leafy spurge patches. *Oecologia* **127**: 287-294.

Jourdheuil, P. (1963) Famille des Chrysomelidae, sous-famille des Halticinae. Balachowsky, A. S. (Ed.) *Entomologie appliquée à l'agriculture*. Paris, Masson et Cie.

Kaszab, Z. (1962) Leaf beetles - Chrysomelidae. Székessy, V. (Ed.), *Magyarorszag Allatvilaga* (*Fauna Hungariae Vol. IX)*. Budapest, Akadémiai Kiadó.

Khan, A. Z., Amad, I., Shaheen, S., Hussain, K., Hafeez, F., Farooq, M. & Noor Ul Ayan, H. (2017) Genetic barcoding and phylogenetic analysis of dusky cotton bug (*Oxycarenus hyalinipennis*) using mitochondrial cytochrome c oxidase I gene. *Cellular and Molecular Biology (Noisy-Le-Grand)* **63**: 59-63.

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

Knodel, J. J., Atkinson, L., Hanson, B., Henson, B. & Olson, D. (2004) Control of Crucifer Flea Beetle in Canola through Insecticide Strategies and Canola Varieties and Forecasting Spring Infestation Risks. [Text article]. Retrieved from https://www.ag.ndsu.edu/CarringtonREC/documents/agronomyrd/docs2004/04%20Canola%20FB.pdf/view

Kolasa, M., Montagna, M., Mereghetti, V., Kubisz, D., Mazur, M. A. & Kajtoch, L. (2017) Preliminary evidence of the horizontal transmission of *Wolbachia* between *Crioceris* leaf beetles (Coleoptera: Chrysomelidae) and their *Asparagus* host plant. *European Journal of Entomology* **114**: 446-454.

Konstantinov, A. S. & Vandenberg, N. J. (1996) Handbook of Palearctic Flea Beetles (Coleoptera: Chrysomelidae: Alticinae). *Contributions on Entomology, International* **1**: 237-439.

Konstantinov, A. S. (1998) On the Structure of the Female Genitalia in Flea Beetles (Coleoptera: Chrysomelidae: Alticinae). *Proceedings of the Entomological Society of Washington* **100**: 353-360.

Konstantinov, A. S., Volkovitsh, M. & Cristofaro, M. (2001) New data on Palearctic *Apthona* (Coleoptera: Chrysomelidae) with description of a new species: Taxonomic and faunistic results of biological control exploration. *Entomological News* **112:** 31-41.

Konstantinov, A. S., Baselga, A., Grebennikov, V. V., Prena, J. & Lingafelter, S. W. (2011) Revision of The Palearctic *Chaetocnema* Species (Coleoptera: Chrysomelidae: Galerucinae: Alticini). Sofia, Bulgaria: Pensoft.

Lamb, R. J. & Turnock, W. J. (1982) Economics of insecticidal control of flea beetles (Coleoptera: Chrysomelidae) attacking rape in Canada. *Canadian Entomology* **114**: 827-840.

Lesage, L. & Majka, C. G. (2010) Introduced leaf beetles of the Maritime Provinces, 9: *Chaetocnema concinna* (Marsham, 1802) (Coleoptera: Chrysomelidae). *Zootaxa* **2610:** 27-49.

Magoga, G., Sassi, D., Daccordi, M., Leonardi, C., Mirzaei, M., Regalin, R., Lozzia, G. & Montagna, M. (2016) Barcoding Chrysomelidae: a resource for taxonomy and biodiversity conservation in the Mediterranean Region. *ZooKeys* **597:** 27-38.

Mänd, M. (2014) Flea beetle (Chrysomelidae: Alticinae) species composition and abundance in different cruciferous oilseed crops and the potential for a trap crop system. *Acta Agriculturae Scandinavica, Section B Soil & Plant Science* **64:** 572-582.

Massimino Cocuzza, G. E. & Cavalieri, V. (2014) Identification of aphids of *Aphis frangulae*-group living on Lamiaceae species through DNA barcode. *Molecular Ecology Resources* **14:** 447-457.

Meier, R., Shiyang, K., Vaidya, G. & Ng, P. K. L. (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology* **55**: 715-28.

Mereghetti, V., Chouaia, B., Limonta, L., Locatelli, D. P. & Montagna, M. (2017) Evidence for a conserved microbiota across the different developmental stages of *Plodia interpunctella*. *Insect Science* **18:** 2450.

Metspalu, L., Kruus, E., Ploomi, A., Williams I. H., Hiiesaar, K., Jõgar, K., Veromann, E. & Mänd, M. (2014) Flea beetle (Chrysomelidae: Alticinae) species composition and abundance in different cruciferous oilseed crops and the potential for a trap crop system. *Acta Agriculturae Scandinavica, Section B Soil & Plant Science* **64**: 572-582.

Montagna, M., Sassi, D. & Giorgi, A. (2013) *Pachybrachis holerorum* (Coleoptera: Chrysomelidae: Chryptocephalinae), a new species from the Appennines, Italy, identified by integration of morphological and molecular data. *Zootaxa* **3741**: 243-253.

Montagna, M., Mereghetti, V., Lencioni, V. & Rossaro, B. (2016) Integrated Taxonomy and DNA Barcoding of Alpine Midges (Diptera: Chironomidae). *PLoS ONE* **11:** e0149673.

Montagna, M., Kubisz, D., Mazur, M. A., Magoga, G., Ścibior, R. & Kajtoch, L. (2017) Exploring species-level taxonomy in the *Cryptocephalus flavipes* species complex (Coleoptera: Chrysomelidae). *Zoological Journal of the Linnean Society* **179:** 92-109.

Nei, M. & Kumar, S. (2000) *Molecular Evolution and Phylogenetics*. New York: Oxford University Press.

Neves, E. F. (1983) On the insect fauna of sugarbeet in Portugal. *Boletim da sociedade portuguesa de Entomologia* **2:** 77-94.

Özdikmen, H., Mercan, N., Cihan, N., Kaya, G., Topcu, N. N. & Kavak, M. (2014) The importance of superfamily Chrysomeloidea for Turkish biodiversity (Coleoptera). *Munis Entomology & Zoology* **9:** 17-45.

Pataki, E. (1967) Der Rübenerdfloh (*Chaetocnema tibialis* Ill.), einer der gefährlichsten Keimschädlinge der Zuckerrübe in Ungarn. *Zeitschrift für Angewandte Entomologie* **59:** 239-248.

Radcliffe, E. B., Hutchinson, W. D. & Cancelado, R. E. (2009) Integrated Pest Management: Concepts, Tactics, Strategies and Case Studies. *Crop Protection* **28:** 453.

Ratnasingham, S. & Hebert, P. D. N. (2013) A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. *PLoS ONE* **8:** e66213.

Sáringer, G. (1990) Flea beetles (Halticinae spp). Jermy, T., Balázs K. (Eds.), *Handbook of Plant Protection Zoology 3/A* (pp. 296-307). Budapest: Akadémiai Kiadó.

Schäffer, S., Zachos, F. E. & Koblmüller, S. (2017) Opening the treasure chest: A DNA-barcoding primer set for most higher taxa of Central European birds and mammals from museum collections. *PLoS ONE* **12:** e0174449.

Scherer, G. (1971) Das Genus *Livolia* Jacoby und seine umstrittene Stellung im System. Eine taxonomische-zoogeographische-evolutionistische Studie (Coleoptera - Chrysomelidae - Alticinae). *Entomologische Arbeiten aus dem Museum G Frey Tutzing bei Muenchen* **22:** 1-37.

Seeno, T. N. & Wilcox, J. A. (1982) Leaf beetle genera (Coleoptera, Chrysomelidae), *Entomography* **1:** 1-221.

Šigut, M., Kostovčík, M., Šigutová, H., Hulcr, J., Drozd, P. & Hrček, J. (2017) Performance of DNA metabarcoding, standard barcoding, and morphological approach in the identification of host-parasitoid interactions. *PLoS ONE* **12:** e0187803.

Soroka, J. J., Bartelt, R. J., Zilkowski, B. W. & Cossé, A. A. (2005) Responses of flea beetle *Phyllotreta cruciferae* to synthetic aggregation pheromone components and plant host volatiles. *Journal of Chemical Ecology* **31:** 1829-1843.

Spilák, K., Garamvölgyi, I. & Balázs, K. (1998) Plant protection of cauliflower and broccoli. *Növényvédelem* **34:** 137-142.

Stef, N. & Buzinovschi, C. (1982) Phytosanitary state of the principal field crops in the Bihor district. *Probleme de Protectia Plantelor* **10:** 261-267.

Stobbs, L. W., Cerkauskas, R. F., Lowery, T. & VanDriel, L. (1998) Occurrence of Turnip Yellow Mosaic Virus on Oriental Cruciferous Vegetables in Southern Ontario, Canada. *Plant Disease* **82**: 351.

Szénási, Á. & Markó, V. (2015) Flea beetles (Coleoptera: Chrysomelidae, Alticinae) in Bt- (MON810) and near isogenic maize stands: Species composition and activity densities in Hungarian fields. *Crop Protection* **77**: 38-44.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**: 2725-2729.

Tangtrakulwanich, K., Reddy, G. V. P., Wu, S., Miller, J. H., Ophus, V. L. & Prewett, J. (2014) Developing nominal threshold levels for *Phyllotreta cruciferae* (Coleoptera: Chrysomelidae) damage on canola in Montana, USA. *Crop Protection* **66:** 8-13.

Tansey, J., Dosdall, L. M. & Keddie, B. A. (2009) *Phyllotreta cruciferae* and *Phyllotreta striolata* responses to insecticidal seed treatments with different models of action. *Journal of Applied Entomology* **133:** 201-209.

Thakur, S. S. & Kashyap, N. P. (1994) New record of *Chaetocnema basalis* Baly (Chrysomelidae: Coleoptera) as a pest of seed potato in Salooni. *Himachal Journal of Agricultural Research* **20:** 97.

Toreniyazov, E. (1999) The main pests of sugarbeet in the conditions of Karakalpakstan. *Sakharnaya Svekla* **3:** 19-20.

Ulmer, B. J. & Dosdall, L. M. (2006) Emergence of overwintered and new generation adults of the crucifer flea beetle, *Phyllotreta cruciferae* (Goeze) (Coleoptera: Chrysomelidae). *Crop Protection* **25:** 23-30.

Verdyck, P. (1998) Genetic differentiation and speciation among four *Phyllotreta* species (Coleoptera: Chrysomelidae). *Biological Journal of the Linnean Society* **64**: 463-476.

Verdyck, P. (1999) Biochemical Systematics of the *Phyllotreta cruciferae* Complex (Coleoptera: Chrysomelidae: Alticinae). *Annals of the Entomological Society of America* **92:** 30-39.

Veromann, E., Tarang, T., Kevväi, R. & Luik, A. (2006) Insect pests and their natural enemies on spring oilseed rape in Estonia: impact of cropping systems. *Agricultural and Food Science* **15:** 61-72.

Vig, K. (1992) Relationships of frequently found *Phyllotreta* species to their hostplants. *Vas Megyei Muzeumok Értesítöje* **20:** 281-294.

Vörös, G. & Garamvölgyi, I. (1998) Plant protection of field-grown kohlrabi. *Növényvédelem* **34:** 193-200.
Yildirim, E. & Özbek, H. (1992) Erzurum Şeker Pancarı Fabrikasına bağlı şeker pancarı üretim alanlarındaki zararlı ve yararlı böcek türleri. *Proceeding of Entomology Congress*, 28-31 January 1992, Adana, 621-635.
Wang, J., Zhang, L., Zhang, Q., Zhou, M., Wang, X., Yang, X. & Yuan, M. (2017) Comparative mitogenomic analysis of mired bugs (Hemiptera: Miridae) and evaluation of potential DNA barcoding markers. *Peer Journal* **5:** e3661.
Warchalowski, A. (2010) *The Palaearctic Chrysomelidae. Identification Keys 2*. Warszawa: Natura Optima Dux Foundation.
Wensing, A., Zimmermann, A. & Geider, K. (2010) Identification of the corn pathogen *Pantoea stewartia* by mass spectrometry of whole-cell extracts and its detection with novel PCR Primers. *Applied and Environmental Microbiology* **76**: 6248-6256.
Zhang, T., Wang, Y.-J., Guo, W., Luo, D., Wu, Y., Kucerova, Z., Stejskal, V., Opit, G., Cao, Y., Li, F.-J. & Li, Z.-H. (2016) DNA barcoding, species-specific PCR and real-time PCR techniques for identification of six *Tribolium* pests of stored products. *Scientific Reports* **6:** 28494.

2.2.4 Appendix

Supplementary materials available at https://onlinelibrary.wiley.com/doi/full/10.1111/jen.12566

2.2.5 Personal contribution to the work

Coinceving of the study, bioinformatics analyses and writing of the manuscript in collaboration with M.M. and D.C.S.

2.3 Research article

Morphology, genetics and Wolbachia endosymbionts support distinctiveness of Monochamus sartor sartor and M. s. urussovii (Coleoptera: Cerambycidae)

Plewa, R., Sikora, K., Gutowski, J. M., Jaworski, T., Tarwacki, G., Tkaczyk, M., Rossa, R., Hilszczanski, J., Magoga, G. & Kajtoch, Ł. (2018) Morphology, genetics and *Wolbachia* endosymbionts support distinctiveness of *Monochamus sartor sartor* and *M. s. urussovii* (Coleoptera: Cerambycidae). *Arthropod Systematics & Phylogeny* **76:** 123-135.

2.3.1 Summary

Monochamus sartor sartor from Central European mountain ranges and *M. s. urussovii* from the Eurasian boreal zone are subspecies whose taxonomic statuses have been questioned. This sawyer beetle is a natural element of spruce forests but is considered to be a timber pest in spruce plantations. In this study, different sets of data (morphology, genetics and ecology) were used to verify the taxonomic status of *M. sartor* across its entire range. Morphologically, not only both subspecies but also European and Asian populations of *M. s. urussovii* were found to be distinct. Genetic data also showed that both subspecies have distinct mitochondrial haplogroups; however, divergence between them is very weak (of ca. 1%), suggesting they split very recently, possibly at the end of the Pleistocene glaciations. Species delimitation methods gave discordant results, either rejecting the species status of *M. s. sartor* and *M. s. urussovii* (Poisson tree processes) or confirming them as distinct taxa (the multispecies coalescent model for species validation). Host plant preferences also partially differentiate the subspecies, as *M. s. urussovii* has a broader diet than the generally monophagous, spruce-dependent *M. s. sartor*. Moreover, each subspecies is infected by different strains of the intracellular bacterium *Wolbachia*, which could be one of the factors causing their genetic isolation, regardless of geographic isolation. Aside from broadening the basic knowledge on the taxonomy and genetics of *Monochamus sartor*, this study shows that any research on these sawyers needs to consider their separate phylogenetic lineages, as do any plans for population management or conservation.

2.3.2 Manuscript

Introduction

Subspecies is the only recognized rank below species level that can receive a name in the zoological code (International Code of Zoological Nomenclature 2000). Recognising subspecies is difficult. According to the biological species concept, organisms belonging to different species can be verified on the basis of their ability to interbreed and produce fertile offspring (Wilson & Brown, 1953; Ehrlich, 1957; Mayr, 1982); however, there is no strict concept for the subspecies level. Biologists can identify subspecies on the basis of whether geographically separate populations of a species exhibit recognizable phenotypic differences (Wilson & Brown, 1953; Ehrlich, 1957; Mayr, 1982). These differences should also be visible in their genotypes, e.g. by the occurrence of distinct phylogenetic lineages. However, the border between intraspecific (interpopulation) diversity and divergence between subspecies is not clearly demarcated. Moreover, distinct phylogenetic lineages can be assigned as Evolutionary Significant Units (Ryder, 1986; Moritz, 1994), which are utilized in conservation genetics but which can also be valuable in taxonomic/phylogenetic studies. In nature, subspecies are mainly unable to interbreed due to geographic isolation of distant populations (Barrowclough, 1982; Cracraft, 1983). Moreover, it is expected that differences between subspecies should usually be less distinct than differences between species, but this assumption is rarely studied with respect to different characters. Most subspecies have been recognized on the basis of slight but significant differences in their morphological features (Nei, 1972; Ball & Avise, 1992), which are identified arbitrarily by observers/taxonomists. Relatively rarely is their taxonomic distinctiveness later verified with other data, e.g. on their genetics or ecologies (Phillimore & Owens, 2006). This especially concerns insects, which are the most diverse group of organisms on Earth (May, 1992; Mora et al., 2011), and particularly beetles (Coleoptera), which form the most species rich order of insects (Farrell, 1998; Grove & Stork, 2000). Detailed, integrative studies that use morphometrics, molecular markers and/or ecological features often find intraspecific variation. The taxonomic status of such polytypic species should be verified to broaden basic taxonomic knowledge, understand phylogenetic relations among units below species level and properly organize studies (to be sure which and how many units or taxa are investigated), as well to conserve/manage rare or economically important taxa.

Here, we focus on the sawyer beetle *Monochamus sartor* (Fabricius, 1787), which belongs to the longhorn beetles (Cerambycidae: Lamiinae). There are about 140 species and 25 subspecies of *Monochamus* worldwide, 50 of which inhabit the Palaearctic – mainly the boreal zone and mountain areas (Danilevsky, 2017; http://insecta.pro/search?search=Monochamus). These species are highly dependent on the dead wood of mainly conifer trees (pines *Pinus* spp., spruces *Picea*

spp., firs *Abies* spp., larches *Larix* spp. and cedars *Cedrus* spp.) (Hellrigl, 1970; Isaev et al., 1988; Wallin et al., 2013). Larvae bore holes inside the wood of thicker branches and trunks, and thus are considered timber pests. Furthermore, species in the genus *Monochamus* are the main vectors of the pinewood nematode, *Bursaphelenchus xylophilus* (Steiner & Buhrer) Nickle (Linit et al., 1983), a quarantine species that causes pine wilt disease (PWD), mainly in East Asia and recently also in Portugal (Kondo et al., 1986; Miller et al., 2013). For this reason, they are considered serious pests of conifer tree plantations (Hellrigl, 1970; Evans et al., 2004; Wang, 2017). On the other hand, they are one of the species responsible for the decay of dead wood in mature conifer forests in the boreal zone and on mountain ranges and are an important food source for numerous bird species (e.g. woodpeckers; Winkler et al., 1995).

Monochamus sawyers have been the objects of numerous taxonomic, systematic and phylogenetic studies (Hellrigl, 1970; Tomminen & Leppänen, 1991; Cesari et al., 2004; Koutroumpa et al., 2013; Wallin et al., 2013; Rossa et al., 2016). Much less is understood on the phylogeography of particular taxa as almost all such studies concern Asian *Monochamus alternatus* Hope (Kawai et al., 2006; Shoda-Kagaya et al., 2008; Hu et al., 2013); just recently a single study on the European *Monochamus galloprovincialis* (Olivier) was published (Haran et al., 2017). On the other hand, such knowledge on the *Monochamus sartor*-complex is insufficient. Until relatively recently, the two currently recognized subspecies, i.e. *M. sartor sartor* and *M. sartor urussovii* (Fischer von Waldheim, 1806), were considered distinct species, i.e. *M. sartor* and *M. urussovii*, respectively (Bense, 1995; Sama, 2002; Löbl & Smetana, 2010). Sláma (2006) used the subspecies rank for both taxa; however, his justification for this splitting is lacking. Some other literature also supports this division (e.g. Wallin et al., 2013), based on the detailed characteristics of adults and the genital morphology of males and females. However, there are still many uncertainties, especially related to the distribution of both taxa. There is no certainty if the subspecific rank of these taxa is appropriate, and if it is, the question arises whether there should be another subspecies distinguished for the populations of north-east Europe.

Both, *M. sartor sartor* and *M. s. urussovii*, occur throughout the natural range of Norway spruce, *Picea abies* (L.) H. Karst, while also rarely utilizing pines and firs. Only in eastern Siberia, Korea, and Japan they are also reported to develop on other spruce species as well as on cedars and birches *Betula* L. sp. (Cherepanov, 1983). However, data on the distribution of these taxa in Europe is often insufficient and sometimes contradictory. For example, Danilevsky (2012) stated that both *M. sartor sartor* and *M. s. urussovii* co-occur in several European countries (e.g. Belarus, Estonia, Latvia, Lithuania and Ukraine). Nevertheless, he questioned the occurrence of the latter subspecies in north-eastern Poland, and further stated that the populations of *M. sartor* from this area are

identical to those found in the Carpathians (i.e. to *M. s. sartor*). As a consequence, the western parts of Belarus (i.e. the Białowieża Primeval Forest) would supposedly be populated by *M. s. sartor*, while the eastern parts by *M. s. urussovii* (Danilevsky, 2012). Meanwhile, a different distribution pattern for both taxa was proposed by Löbl & Smetana (2010) and later Wallin et al. (2013), who suggested the two taxa co-occur in six European countries, namely in Belarus, Czechia, Latvia, Lithuania, Estonia and Poland. These authors, however, did not specify the areas of distribution of *M. s. urussovii* in these countries.

The distributional pattern is less obscure for *M. s. sartor*, which is characterized by a European type of distribution – its range is less extensive than *M. s. urussovii* and covers the Alps, Carpathians, Dinaric Alps and Bulgarian mountains, and their foothills (Heyrovský, 1955; Mikšič & Korpič, 1985; Dominik & Starzyk, 2004; orig.inf.).

Morphology and morphometric information complement molecular data. Moreover, these types of data are often supplemented with ecological features (e.g. habitat requirements and interactions with other organisms, such as host plants for herbivores), and together such a comprehensive elaboration of species distinctiveness and relations is called "integrative taxonomy" (Dayrat, 2005; Schlick-Steiner et al., 2010). Among ecological features, relations with symbionts or parasites are often utilized (Valentini et al., 2009), with great emphasis on microbiota (Steinert et al., 2000; Hosokawa et al., 2006). Intracellular bacteria can be especially important for arthropods, as some endoparasites or endosymbionts can directly influence host fitness, development and diversity, which may in turn have implications on host speciation (Hurst & Jiggins, 2000; Engelstädter & Hurst, 2009). Notable examples of such endosymbionts/parasites are the maternally inherited bacteria *Wolbachia* (Bourtzis & O'Neill, 1998; Stouthamer et al., 1999; Zchori-Fein & Perlman, 2004; Kikuchi, 2009; White et al., 2009, 2011). Finally, phylogenetic and systematic studies are especially challenging in cases where the examined taxa are presumed to be of subspecies status (Avise & Wollenberg, 1997; Presgraves, 2010). This is because often it is hard to decide to which (intraspecific, taxonomic) level these taxa should be assigned.

Here, we use a combination of morphological and molecular (including *Wolbachia* endosymbiont diversity) features to identify phylogenetic lineages within populations of *Monochamus sartor* sawyers and particularly to examine the taxonomic statuses of presumed subspecies. We aimed to verify the following hypotheses: that i) *Monochamus sartor sartor* from mountainous areas of Europe (i.e. the Carpathians and Alps) and *M. s. urussovii* from the boreal zone of Eurasia are distinct subspecies, which evolved in distinct Pleistocene refugia; ii) *M. s. urussovii* from its westernmost range in centraleastern and northern Europe is taxonomically distinct from *M. s. urussovii* in northern and eastern Asia; iii) *Wolbachia* infection differs between *M. s. sartor* and

M. s. urussovii, which suggests it played a role in the divergence of these sawyers.

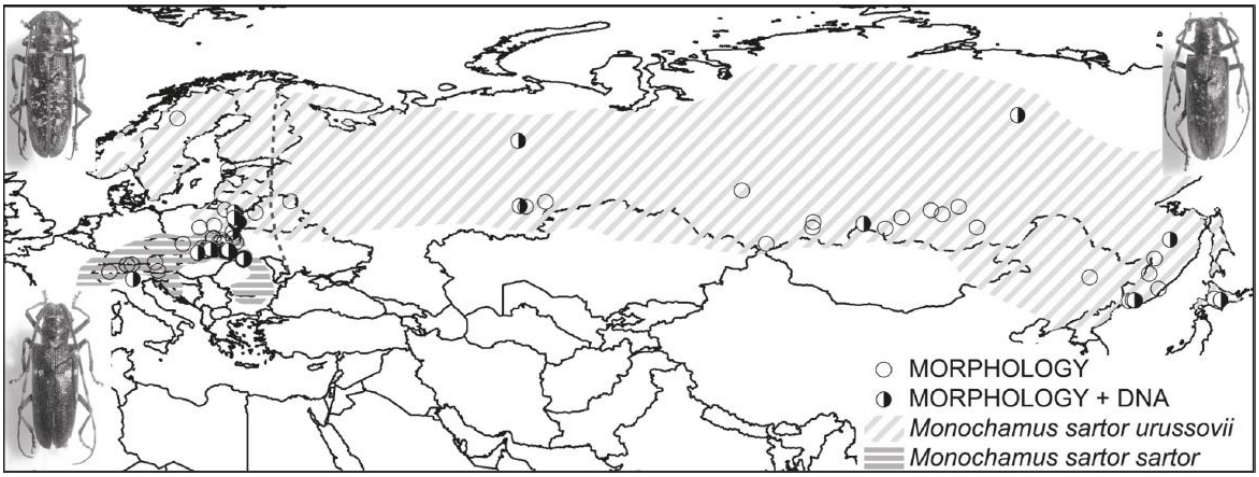

Figure 1. Distribution of *Monochamus sartor sartor* and *M. s. urussovii* in Eurasia with localization of sampling sites for morphological and molecular studies. Broken line – approximate border between eastern and western populations of *M. s. urussovii*.

Material and methods

Sampling of specimens

All specimens of *Monochamus sartor sartor* and *M. s. urussovii* used in our study were collected between 1902 – 2016 by various entomologists (Fig. 1). Thus, most of them were dried specimens that we borrowed from various institutions and private entomological collections. Furthermore, adult specimens were collected by us in 2014 – 2017 in NE (Białowieża Forest, Augustów Forest and Knyszyn Forest) and SE (Bieszczady, Beskid Niski, and Pieniny Mts.) Poland (Fig. 1). In total, 531 specimens of both subspecies of *M. sartor* were collected (see Supplementary Tables S1 and S2 for detailed characteristics of the gathered material).

Morphological analyses

The morphological study was based on selected body characteristics of 523 adult sawyers (247 females, 276 males). Namely, the maximal width of thorax (i.e. the mesothorax section) and the length of the right elytra of each specimen were measured. The measurements were taken using a Zeiss Stemi 2000-C stereomicroscope, within a 0.1 mm accuracy.

The normality of the distribution of data was checked using a Shapiro-Wilk test. A Kruskal-Wallis test for independent groups, i.e. beetles from: mountainous areas of Europe (group I), boreal Europe (group II), and Central and East Asia (group III), was used to compare the studied morphological features. Due to expected sexual dimorphism both sexes of sawyer beetles were treated separately.

Statistical analyses were carried out in Statistica 10 (StatSoft 2011).

Table 1. Genetic diversity of mitochondrial DNA in examined populations of *Monochamus sartor* sawyers. — Abbreviations: N – sample size, H – number of haplotypes, S – number of segregating sites, Hdiv – haplotype diversity, πdiv – nucleotide diversity, SD – standard deviation.

Subspecies	Area	N	H	S	Hdiv ± SD	πdiv ± SD
Monochamus sartor urussovii	Asia eastern	8	7	11	1.00 ± 0.08	0.003 ± 0.002
	Asia central	5	5	10	1.00 ± 0.13	0.004 ± 0.002
	Asia - total	13	13	21	1.00 ± 0.03	0.005 ± 0.002
	Poland north-eastern	16	13	17	0.98 ± 0.03	0.004 ± 0.002
	Eurasia	29	25	28	0.99 ± 0.01	0.004 ± 0.002
Monochamus sartor sartor	Alps	4	3	2	0.83 ± 0.22	0.001 ± 0.001
	Carpathians	26	10	9	0.70 ± 0.09	0.001 ± 0.001
	Alps & Carpathians	30	12	11	0.71 ± 0.09	0.001 ± 0.001

Molecular analyses

Molecular analyses were conducted on 59 specimens representing all parts of both subspecies' ranges in Europe and Asia (details listed in Table 1). Most specimens were directly preserved in 99% ethanol and kept in –20°C freezer until use – this concerns beetles collected in SE (*M. s. sartor*) and NE Poland (*M. s. urussovii*). Other specimens (mostly *M. s. urussovii* from Russia and Japan) were preserved as dry samples.

Laboratory procedures

DNA was extracted from internal tissues of abdomens (specimens were retained for morphological measurements and collection) using Nucleospin Tissue kits (Macherey-Nagel) following the manufacturer's instructions. Two different beetle genes were amplified, sequenced and used for the following analyses. Partial sequences of mitochondrial cytochrome oxidase subunit I (*cox1*) and nuclear elongation factor 1-alpha (*ef-1α*) were amplified using primers C1-J-1751 and L2-N-3014 (Simon et al., 1994), and EFs149 and EFa1R (Normark et al., 1999; Sanz Muñoz, 2010), respectively. The details of amplification, purification and sequencing procedures were reported in Kubisz et al. (2012). *Cox1* was amplified for all examined beetles, whereas *ef-1α* could only be amplified from fresh-preserved specimens (for *M. s. sartor* these were collected from the Carpathians and for *M. s. urussovii* from North-eastern Poland). Because the initial sequencing of *ef-1α* revealed no polymorphism in *M. s. sartor* and *M. s. urussovii*, we did not analyse this gene further. Moreover, we downloaded all available cox1 sequences of *M. s. sartor*, which originated from Italian (Alpine) specimens (GenBank accession numbers: AY260838 – AY2608340).

Wolbachia infection was initially screened in all individuals with ftsZ_F1 / ftsZ_R1 and hcpA_F1 / hcpA_R1 primers for two *Wolbachia*-specific genes (Baldo et al., 2005). Next, all positively

infected individuals were genotyped with respect to all five genes included in the Multilocus Sequence Typing system accepted for *Wolbachia* (details available at http://pubmlst.org/wolbachia/). We excluded from the analysis all dry-preserved beetle specimens as we could not rule out whether lack of amplification of bacterial genes really indicated a lack of infection or DNA from these specimens was just too degraded.

The sequences of presumed *Monochamus* and *Wolbachia* genes were compared with the online NCBI databank using the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) to check if the primers had specifically amplified the targeted sequences of sawyers and α-proteobacteria.

The obtained electropherograms, after correction using BioEdit v.7.0.5.2 (Hall, 1999), were deposited in GenBank (accession numbers: MF327393 – MF327421 and MF371175 – MF371201 forcox1; MF405509 – MF405514 for ef-1α; MF405515 – MF405520 for gatB, MF405521 – MF405526 for coxA, MF405527 – MF405532 for hcpA, MF405533 – MF405539 for ftsZ and MF405540 – MF405545 for fbpA).

Protein-coding DNA sequences (*cox1* and *ef-1α*) were aligned using MAFFT (Katoh et al., 2005). Pairwise nucleotide divergences for both sawyer markers were calculated using MEGA v6 (Tamura, et al. 2013).

Genetic diversity

Haplotypes were identified and standard genetic indices such as haplotype diversity (Hdiv), nucleotide diversity (πdiv), number of private haplotypes (Hpriv) and number of segregating sites (S) for populations were computed using the program DnaSP v.5 (Librado & Rozas, 2009). Population samples were grouped according to their geographical locations (Table 1). FST indices were calculated using ARLEQUIN v.3.5 (Excoffier & Lischer, 2010). A Mantel test (Mantel, 1967) was performed in the program ARLEQUIN to check if the genetic structure of the sampled localities (five) fits an isolation by distance model (IBD) (Slatkin, 1993), using pairwise FST values and straight-line geographic distances in kilometers. To test for the presence of barriers between populations, an analysis of molecular variance (AMOVA) was conducted in ARLEQUIN. Moreover, a minimum-spanning (MS) haplotype network (Bandelt et al., 1999) was reconstructed forcox1 in PopArt (http://popart.otago.ac.nz/).

Phylogenetic analyses and species delimitation

Nucleotide substitution models were estimated for the datasets using the Smart Model Selection tool implemented in PhyML 3.0 software (Guindon et al., 2010) and the best nucleotide evolution

model was selected according to the Bayesian information criterion. For *cox1*, GTR was selected as the best model of nucleotide evolution both for ingroup and outgroup taxa.

For *cox1* phylogenetic reconstruction, single sequences were randomly selected from each sample site and the following outgroup taxa were added from GenBank: *Monochamus galloprovincialis* (GenBank accession number: AY260835), *M. saltuarius* (Gebler) (AY260842), *M. alternatus* (KF737828), *M. sutor* (L.) (AY264403) and *Anoplophora glabripennis* (Motschulsky) (EU914688). Phylogenetic trees were reconstructed adopting the Bayesian inference (BI) and maximum likelihood (ML) approaches. BI was performed using Mr-Bayes 3.2.2 (Ronquist et al., 2012) in two independent runs, each with one cold and five heated Markov chains ($\lambda = 0.1$) run each, for 2×107 generations that were sampled every 100 generations. Stationarity was considered to be reached when the average standard deviation of the split frequencies was less than 0.01; however, the convergences of each run were also visually inspected using TRACER (Drummond et al., 2012). An appropriate number of sampled trees were discarded as burn-in, and a majority-rule consensus tree was obtained. The ML analyses were performed using PhyML 3.0 software (Guindon et al., 2010) using the command line version. Branch support was obtained by the Approximate Likelihood-Ratio Test (aLRT) (Anisimova & Gascuel, 2006).

Molecular species delimitation analyses were performed oncox1 of all *Monochamus* species included in this study, adopting the tree-based method the Bayesian Poisson tree process model (bPTP; Zhang et al., 2013), and the Bayesian Markov chain Monte Carlo program for Phylogenetic and Phylogeographic analyses under the multispecies coalescent model (BPP; Yang, 2015).

bPTP and BPP methods have been extensively used to recognize and delimit species (e.g. Hambäck et al., 2013; Cranston & Krosh, 2015; Lecocq et al., 2015), as well as to support the description of new insect taxa (e.g. Leaché & Fujita, 2010; Montagna et al., 2016a). bPTP analysis, performed on the BI tree, was carried out with the bPTP web server (http://species.h-its.org/ptp/) with the following parameters: 500,000 MCMC generations, thinning every 200 generations, and 0.2 % of generation discarded as burn-in. The BPP guide tree was drawn on the basis of the BI tree topology. We performed A01 and A11 analyses four times, each with different combinations of prior gamma distributions: i) Θ : G(2,200), τ : G(2,400); ii) Θ : G(2,200), τ : G(2,200); iii) Θ : G(2,200), τ : G(2,2000); and iv) Θ : G(2,2000), τ : G(2,200). Each analysis consisted of 100,000 MCMC generations sampled every 20 generations and discarding the first 20% of the samples as burn-in. Moreover, mean genetic distances among sites were calculated using MEGA5 (Tamura et al., 2011) under the Kimura 2-parameter model (K2P).

Wolbachia infection

Allelic profiles of MLST genes were generated for each infected individual. Next, we utilized an approach similar to that of Montagna et al. (2014) to compare allelic profiles generated from *Monochamus* beetles with some representative sequence types from other species that harbored bacteria belonging to supergroups A (ST-1 from *Drosophila melanogaster* Meigen), B (ST-15 from *Drosophila simulans* Sturtevan), D (ST-35 from unspecified nematode), F (ST-8 from *Cimex lectularius* L.), and H (ST-90 from *Zootermes angusticollis* (Hagen)). Moreover, the allelic profiles found for the only European beetles with full allelic profiles in the MLST database were added to this set of MLST sequences: *Eusomus ovulum* Germar (Mazur et al., 2016), *Oreina cacaliae* (Schrank) (Montagna et al., 2014), and *Crioceris quinquepunctata* Scopoli (Kubisz et al., 2012). We then used the generated alignment of MLST genes for the construction of a phylogenetic network in SplitsTree4 (Huson & Bryant, 2006) by using neighbor-net algorithm distance estimates. In contrast to traditional phylogenetic trees, this allows for visualization of multiple connections among examined sequences, which can represent recombination events. The PHI test implemented in SPLITSTREE v. 4 (Huson & Bryant, 2006) has been shown to identify the presence/absence of recombination within a range of sequence samples (both insect and bacterial markers) with a low false-positive rate (Bruen et al., 2006). The PHI test rejected the hypothesis assuming recombination among MLST genes ($p = 1.000$). Additionally, the most similar hits to all MLTS (gene) sequences generated from *Monochamus* sawyers were identified with the BLAST search tool (Altschul et al., 1990) against NCBI GenBank resources.

Results

Morphological differentiation

Statistical analyses revealed significant differences between the studied size features of *M. sartor* populations from the three areas of distribution (female thoracic width: $H = 53.78$, $df = 2$, $P < 0.001$; female elytral length: $H = 42.13$, $df = 2$, $P < 0.001$; male thoracic width: $H = 24.03$, $df = 2$, $P < 0.001$; male elytral length: $H = 16.74$, $df = 2$, $P < 0.001$). In general, females from boreal Europe (group II) had a smaller body size, as reflected by mean thoracic width and elytral length, than females from both mountainous areas of Europe and Central and East Asia (groups I and III, respectively). A similar pattern was observed for adult males, in which the studied body characteristics of specimens from boreal Europe differed significantly from those from mountainous areas of Europe, but were not different from specimens from Central and East Asia (Fig. 2).

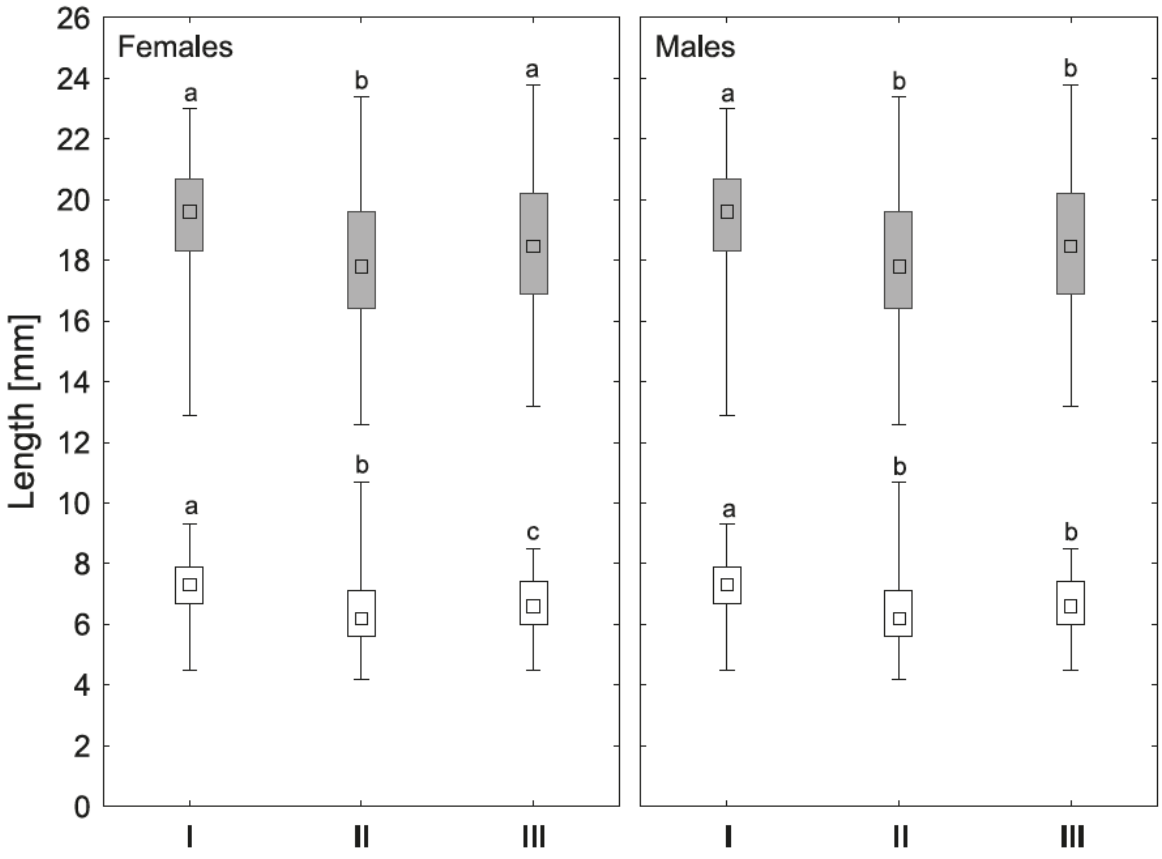

Figure 2. Differences between the length of the right elytra (gray boxes) and the width of thorax (white boxes) of female and male *Monochamus sartor* sawyers from three distinguished areas of distribution: I – montane areas of Central Europe (Carpathians and Alps), II – boreal Europe, and III – Central and East Asia. Squares indicate medians, boxes indicate 25th and 75th percentiles and whiskers indicate minimum and maximum values; different letters indicate significant differences between studied populations; $p < 0.05$.

Molecular differentiation

Genetic diversity

After trimming of ambiguous fragments of *ef-1α* sequences, the final alignment was 600 bp long. There were no stop codons, and only one indel (3 bp) differentiated *A. glabripennis* from *Monochamus* species. Due to lack of polymorphism in ef-1α in both *M. sartor* subspecies, all below-mentioned analyses were based only on the *cox1* dataset. The *cox1* alignment was 1187 bp long, and no stop codons or indels were detected. Genetic diversity was high in *M. s. urussovii* and was similarly high across all its geographic groups of population samples (Table 1). On the other hand, *M. s. sartor* had much lower genetic diversity (Table 1).

Monochamus sartor shows weak but significant isolation by distance (Mantel test: R = 0.245, P = 0.003). AMOVA showed that 66.69% of the molecular variance could be attributed among groups of population samples, 24.80% within populations and only 8.51% among population samples

within groups (FSC = 0.255, FST = 0.751, FCT = 0.666, all P < 0.001). FST values were low between Asian and Polish populations of *M. s. urussovii* (0.056) and very high between *M. s. urussovii* and *M. s. sartor* (0.770 and 0.783, respectively).

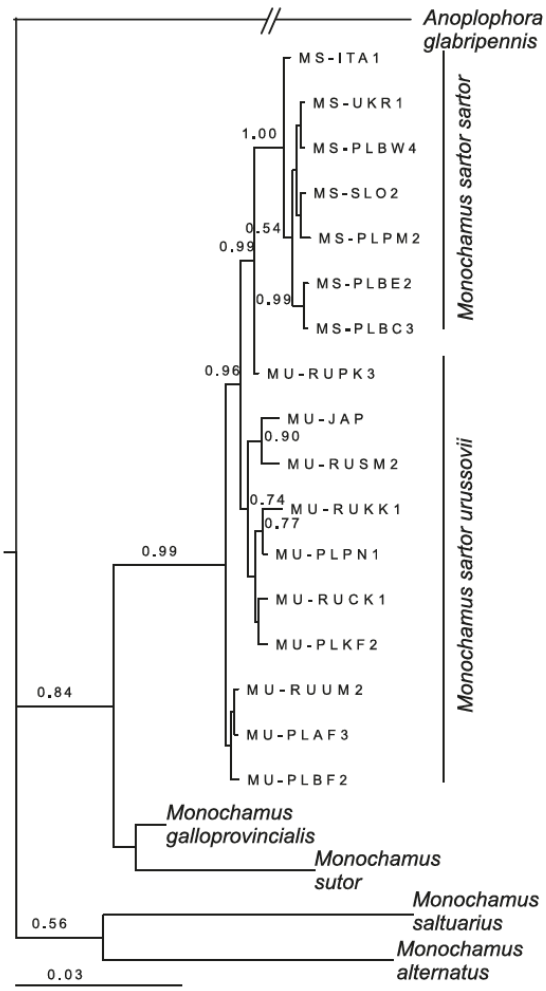

Figure 3. Bayesian phylogenetic tree reconstructed for examined *Monochamus sartor sartor* and *M. s. urussovii* sawyers on the basis of polymorphism of *cox1* gene. Values indicate support of branches (posterior probabilities).

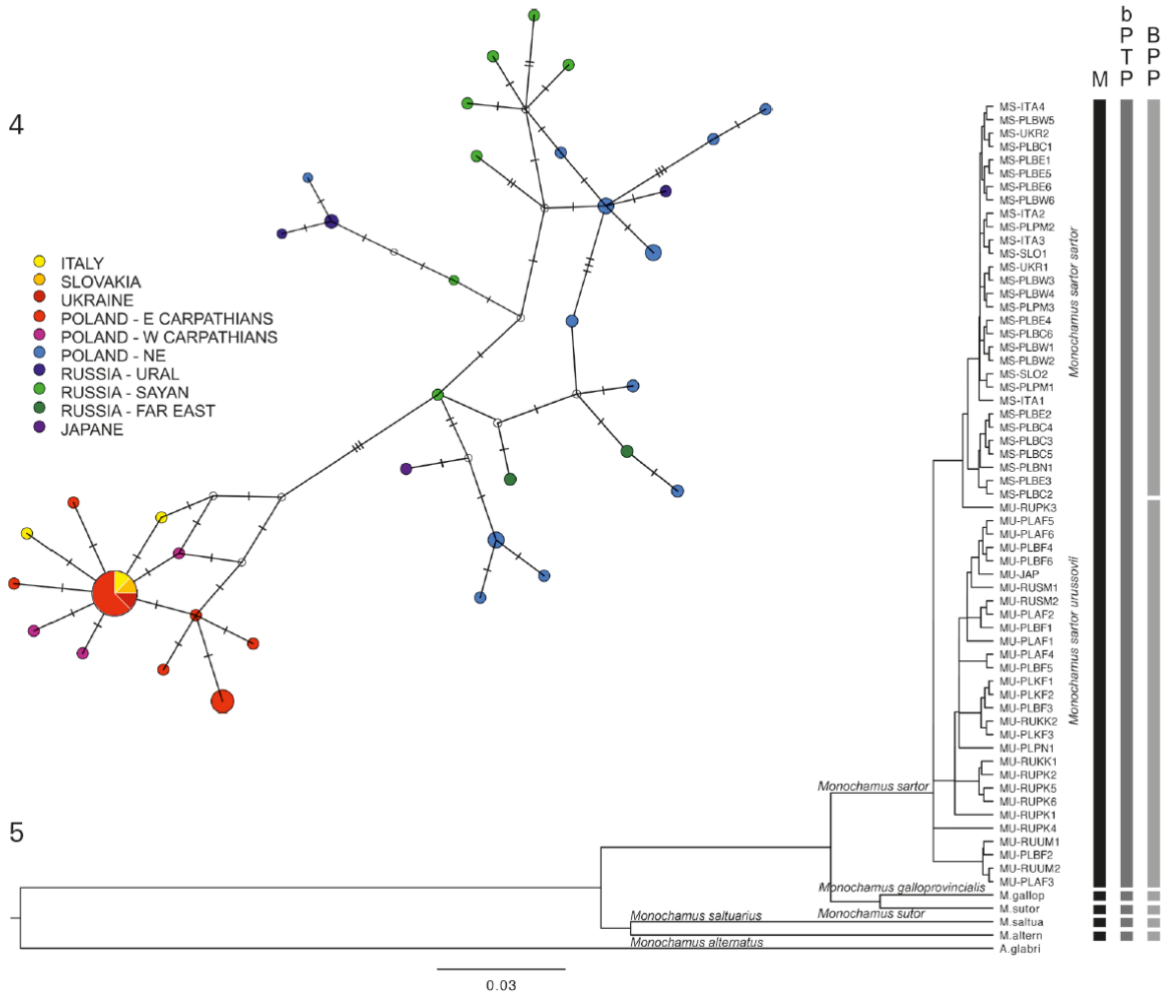

Figure 4. Minimum-spanning haplotype network of *cox1* gene sequenced for *Monochamus sartor sartor* and *M. s. urussovii* sawyers.

Figure 5. Ultrametric Bayesian phylogenetic tree reporting the results of the species delimitation analyses. Vertical bars correspond to morphology (M; black) and to the species delimitation results obtained by bPTP and BPP methods, respectively in dark and light grey.

Phylogenetic analyses and species delimitation

Both BI and ML methods resulted in trees of congruent topologies and therefore only BI trees were presented. Phylogenetic reconstruction calculated for both markers supported the monophyly of *M. sartor*. *Ef-1α* failed to distinguish *M. s. urussovii* from *M. s. sartor*, as both taxa shared a haplotype of this gene. On the other hand, *cox1* suggested the presence of two clusters: *M. s. urussovii* and *M. s. sartor*; however, they were not monophyletic, as the second was nested within the first (Fig. 3).

The haplotype network showed that *M. s. urussovii* and *M. s. sartor* form distinct clusters, whose closest haplotypes are divided by only seven substitutions (Fig. 4). Within *M. s. urussovii* K2P nucleotide divergence was in the range of 0.1 – 0.7%, within *M. s. sartor* 0.0 – 0.2%, and between both subspecies 0.4 – 1.2%.

Species delimitation with the bPTP method recognized five entities (Fig. 5; 95% CI 5 – 7 entities) with a Bayesian posterior probability ranging from 0.94 to 1. The method supported the species distinctiveness of all *Monochamus* species including *M. sartor*, but rejected the distinctiveness of two subspecies of the latter. Whereas, BPP analyses adopting different priors were in close agreement on the best tree topology (i.e. (*M. alternatus*, (*M. saltuarius*, ((*M. galloprovincialis*, *M. sutor*), (*M. sartor urussovii*, *M. sartor sartor*))))) and in accordance in recognizing the presence of 6 entities, with the two *M. sartor* subspecies delimited as separate species (delimitation posterior probability ranging from 0.84 to 0.99) (Fig. 5).

Wolbachia infection

Both of the examined subspecies of *Monochamus sartor* (from the Carpathians, n = 26 and NE Poland, n = 16) were found to be infected and all tested specimens harboured *Wolbachia*. Both subspecies were infected with two *Wolbachia* strains, but each harbored different strains, so overall the species was found to be infected by four strains – all belonging to the A supergroup (Fig. 5). *Monochamus sartor urussovii* was infected by strain 1 in Białowieża and Knyszyn Forests and by strain 2 in Augustów Forest (Fig. 6). *Monochamus sartor sartor* was infected by strains 3 and 4 (Fig. 5). According to a BLAST search against the MLST database, the most similar loci were found in the following species: *Evagetes parvus* (Cresson) wasp (3 genes similar), *Ceutorhynchus obstrictus* (Marsham) weevil (2 genes similar) and *Agelenopsis naevia* (Walckenaer) spider (3 genes similar). The BLAST comparison of *Wolbachia* genes against GenBank resources showed that similar variants were found in the following species: *Ceutorhynchus obstrictus* weevil, *Oreina liturata* (Scopoli), *Altica impressicollis* (Reiche) and *Hermaeophaga mercurialis* (Fabricius) leaf beetles, *Leptopilina clavipes* (Hartig) wasp, and *Lutzomyia stewarti* (Mangabeira & Galindo) fly.

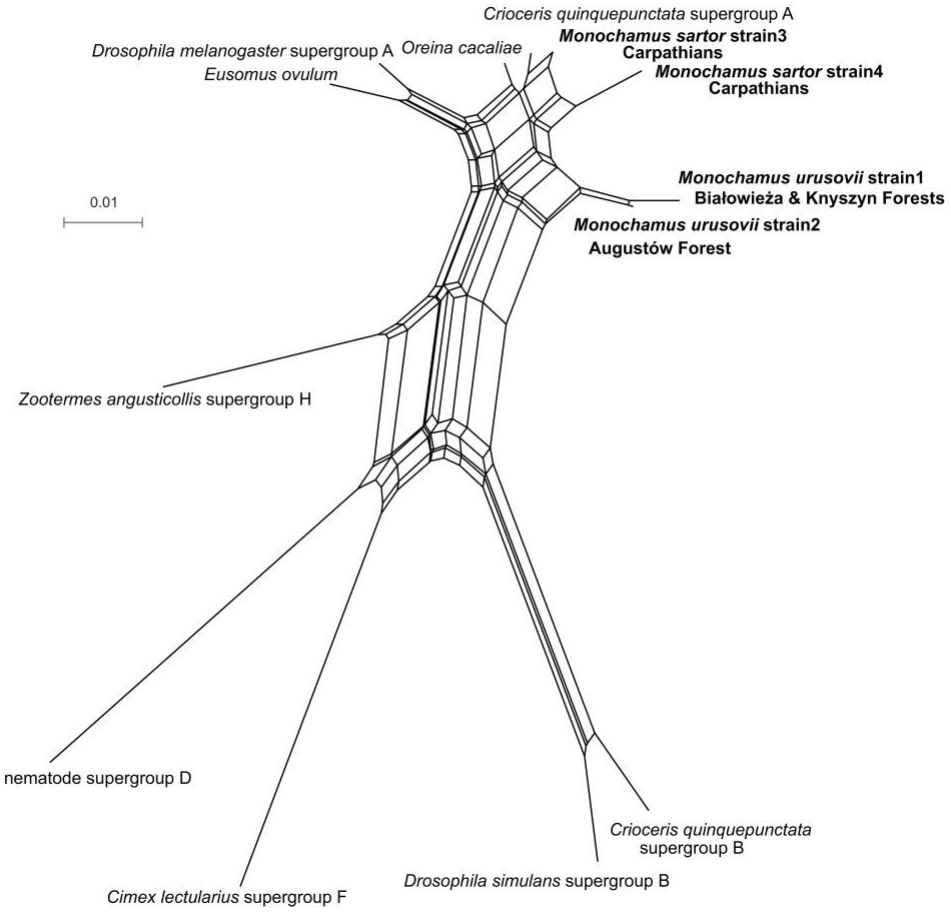

Figure 6. Median network reconstructed for *Wolbachia* strains generated from examined *Monochamus sartor sartor* and *M. s. urussovii* sawyers on the basis of polymorphism of five *Wolbachia* genes used for Multilocus Sequence Typing (*gatB, coxA, hcpA, ftsZ, fbpA*). *Wolbachia* strains from examined sawyers are shown with relations to the representative strains of all *Wolbachia* supergroups found in Multilocus Sequence Typing databases and some strains found in other European beetles.

Discussion

This study aimed to verify whether *Monochamus sartor sartor* and *M. s. urussovii* are valid species or just subspecies. This issue has been the topic of many studies that often show contradictory conclusions (Hellrigl, 1970; Isaev et al., 1988; Bense, 1995). The preliminary phylogenetic study of Cesari et al. (2004) rejected the distinctiveness of these subspecies, however only on the basis of very limited sampling. On the other hand, Rossa et al. (2016) showed that these two subspecies differ significantly with respect to wing venation. Moreover, the host plant preferences of both subspecies (*M. s. sartor* is associated almost exclusively with spruces, whereas *M. s. urussovii* develops on a larger variety of hosts, especially in its Asiatic range) show that there are some ecological and possibly adaptational differences between these taxa – however, these differences

could also have arisen due to phenotypic plasticity. In this study, all collected types of data (morphology, genetics and ecology) support the distinctiveness of both taxa, while simultaneously showing that their differentiation is very shallow. These integrative data suggest that species status should not be supported. On the other hand, *Monochamus sartor* could be an example of recent divergence, with a split forming between the boreal *M. s. urussovii* and mountain *M. s. sartor*. The shallow divergence and *M. s. sartor* haplotypes nested within *M. s. urussovii* are arguments against treating these two subspecies as distinct species. But, as sister species often do not form reciprocal monophyletic clades in molecular data (Knowles & Carstens, 2007), we are not able to definitively rule out that these two subspecies are in fact separate species. This issue requires some further study, like experimental crossing of members of both groups to verify if they produce offspring. If yes, the genetics, reproduction, fitness, ecology and behaviour of progeny should be examined to check if there are any postzygotic barriers supporting species status.

There are several morphological characters that distinguish *M. s. sartor* from the Central-European mountains from *M. s. urussovii* from the Eurasian semiboreal and boreal zones. These characters include density of hairs in distal parts of elytra and their punctuation (e.g. Plavil'shhikov, 1958; Wallin et al., 2013). The question is whether these differences are just due to phenotypic plasticity and environmental adaptations (Grenier et al., 2016), or in fact represent phenotypic proof for the existence of two separate species. The genetic data collected in this study confirm that mountain and boreal populations of *Monochamus sartor* are characterized by different mitochondrial haplogroups and that no haplotypes are shared between subspecies, but also that haplotypes of *M. s. sartor* are nested within *M. s. urussovii*, according to the phylogenetic tree reconstruction. On the other hand, the closest haplotypes belonging to these two groups are only distant by approximately 1%. This value is much below the threshold that is usually observed between sibling species, which for Cerambycidae is higher than 4% (Nakamine & Takeda, 2008; Ohbayashi & Ogawa, 2009), and similar or larger interspecific distances have been observed for other closely related beetles (e.g. Kubisz et al., 2012; Montagna et al., 2016b). Within *Monochamus* sawyers, interspecific distances between *M. galloprovincialis* and *M. sutor* reach 3.1% (Koutroumpa et al., 2013). The species delimitation methods gave discordant results concerning the status of *M. s. sartor and M. s. urussovii.* Bayesian PTP rejected the species distinctiveness of the two taxa (while simultaneously supporting species status of the other *Monochamus* species), while BPP supported their species status. For *M. s. sartor* and *M. s. urussovii*, it is hard to tell which method gave more reliable results, since the results obtained with bPTP can only be considered putative species that should be confirmed by other methods (Zhang et al., 2013). On the other hand, the coalescence adopted by BPP can only delimit population structure and not species boundaries (Sukumarana & Knowles,

2017). In any case, these contrasting results may highlight limitations associated with the use of single locus data.

Other evidence, from the analyses performed on the available sequence data, suggest that there is no separation between *M. s. sartor* and *M. s. urussovii*. For example, the low genetic distance between them is similar to the distances of 1 – 2% that have been observed between presumed subspecies of *M. galloprovincialis* and *M. sutor*, whose subspecies statuses have also been questioned (Koutroumpa et al., 2013). A lack of genetic support for the distinctiveness of both taxa is also indicated by the presence of the same ef-1α haplotype (the only one) in both *M. s. sartor* and *M. s. urussovii*, but explanation of this low (or lack) of nuclear variation needs further studies with more variable markers like microsatellites. It is possible that *Monochamus sartor* sawyers just followed the recent expansion of its host plant – spruce (Taberlet et al., 1998; Latałowa & van der Knaap, 2006), which is known to be of double (boreal and mountain) origin in some areas (e.g. Białowieża Forest) (Latałowa & van der Knaap, 2006; Dering & Lewandowski, 2009; Tollefsrud et al., 2015; Nowakowska et al., 2017). Moreover, timber harvesting and transportation could also have facilitated passive migrations of sawyers across large distances (Etxebeste et al., 2015).

Another question in our study was to verify whether *M. s. urussovii* from Asia and from its westernmost range in Eastern Europe represent the same or distinct units. In this case, morphology and genetics show somewhat inconsistent patterns. Concerning morphology, there are important differences between these populations in females, but not males. Rossa et al. (2016) showed that wing venation of individuals from these two populations is distinct but only slightly if compared to *M. s. sartor*. They even suggested that *M. s. urussovii* from NE Poland could be hybrids between *M. s. sartor* and the Asian *M. s. urussovii*, but this could also be a result of ongoing gene flow (probably mediated by males if considering only mtDNA distinctiveness of these populations). In contrast to this, mitochondrial DNA did not indicate separation of *M. s. urussovii* populations, as haplotypes from Asia and NE Poland did not form distinct clusters and in some cases the same or highly related haplotypes were found in very distant localities. Such patterns are quite common for boreal species that have wide distributions across the Palaearctic boreal zone and which probably expanded from Asian refugia after the end of the Pleistocene glaciations following the expansion of coniferous forests. The phylogeography of some cambioxylophagous beetles follows this pattern (e.g. some bark beetles, Stauffer et al., 1999; darkling beetles, Painter et al., 2007). Moreover, the presented genetic data cannot solve the hybrid origin of *M. s. urussovii* from NE Poland, as the nuclear marker used in this study was found to be monomorphic across the entire species range. Further studies with microsatellites or single nucleotide polymorphism loci are needed to verify this hypothesis.

History of *Monochamus sartor* probably follows the history of boreal tree species, particularly spruce – its main host plant. The current range of *Monochamus sartor* is strictly associated with the range of *Picea* spp., and rarely has this species been found foraging on other conifers or birches. The genetic diversity of its two subspecies strongly indicates that they survived glacial periods in at least two refugia – in the Alps and/or in the Carpathians (*M. s. sartor*) and most probably somewhere in Asia or in Asia together with Eastern Europe (*M. s. urussovii*). Foothills of both the Alps and Carpathians (especially the Southern Carpathians) are known refugial areas for many species, which survived there unfavorable glacial periods in so-called "cryptic" northern refugia (Steward & Lister, 2001; Schmitt & Varga, 2012). Also, Eastern Europe (southern Russian Plains) and East Asia are known refugial areas for numerous continental and boreal species (Stewart et al., 2010). Low divergence between the two subspecies could suggest that their isolation occurred quite recently – probably during one of the last glacial periods. Worth noting is that during the Holocene gene flow between the two subspecies has probably not occurred or has been restricted to male-mediated dispersal.

Previous studies on the microbiota of some *Monochamus* species either found no *Wolbachia* infection like in *M. galloprovincialis* (Vicente et al., 2013) or showed that although *M. alternatus* (Asian species) is currently not infected, it had to have been in the past as it carries some *Wolbachia* genes in its genome (Aikawa et al., 2009, Aikawa et al., 2014). In this study, for the first time, we have confirmed the presence of *Wolbachia* in *Monochamus* species. Interestingly, we found that both subspecies of *Monochamus sartor* are infected (at least in their examined populations from the Carpathians and NE Poland). Moreover, both subspecies harbor different strains – two each, which all belong to supergroup A but are distinct from each other. The presence of different bacterial strains in the two subspecies could further indicate their distinctiveness. Discussing the role of this bacterium in subspecies formation via isolation (e.g. caused by cytoplasmic incompatibility) would be too speculative without further studies. *Wolbachia* can also be used as a biocontrol agent against some insect pests (Lacey & Goettel, 1995; Zabalou et al., 2004), so studies in this direction could also be interesting for controlling outbreaks of *Monochamus sartor* populations, especially with respect to its role as a vector for the pinewood nematode, *Bursaphelenchus xylophilus* (Linit et al., 1983), a quarantine species that causes PWD (Kondo et al., 1986; Miller et al., 2013).

Conclusions

Previous uncertainty on the taxonomic status of *Monochamus sartor sartor* and *M. s. urussovii*, and the Asian and European populations of the latter have been solved in this study. All the gathered

types of data (morphology, genetics and ecology) indicate that these two subspecies should not be considered valid species, in contrast to what has been proposed in the past (Bense, 1995; Sama, 2002; Löbl & Smetana, 2010). The question is whether these presumed subspecies should be considered subspecies. The data presented in this study provide several forms of evidence that despite weak divergence, the boreal and mountain populations differ with respect to their morphology, diversity of endosymbiotic bacteria and plasticity of host plant use. The evidence supports the hypothesis that they should be considered separate subspecies that split quite recently. Aside from broadening the basic knowledge on the taxonomy and genetics of *Monochamus sartor*, this study shows that any research on these sawyers needs to consider their separate subspecies status. Moreover, any plans for population management (if considering them to be forest pests) or population conservation (if considering them to be natural elements of mature forests with high shares of dead wood) of these longhorn beetles should take into account that there are two groups, which differ with respect to numerous characters and therefore could react in different ways to forest management or conservation practices.

2.3.3 References

Aikawa, T., Anbutsu, H., Nikoh, N., Kikuchi, T., Shibata, F. & Fukatsu, T. (2009) Longicorn beetle that vectors pinewood nematode carries many *Wolbachia* genes on an autosome. *Proceedings Biological Sciences* **276:** 3791-3798.

Aikawa, T., Nikoh, N., Anbutsu, H. & Togashi, K. (2014) Prevalence of laterally transferred *Wolbachia* genes in Japanese pine sawyer, *Monochamus alternatus* (Coleoptera: Cerambycidae). *Applied Entomology and Zoology* **49:** 337-346.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215:** 403-410.

Anisimova, M. & Gascuel, O. (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biology* **55:** 539-552.

Avise, J. C. & Wollenberg, K. (1997) Phylogenetics and the origin of species. *Proceedings of the National Academy of Sciences of the USA* **95:** 7748-7755.

Baldo, L., Lo, N. & Werren, J. H. (2005) Mosaic nature of the *wolbachia* surface protein. *Journal of Bacteriology* **187:** 5406-5418.

Ball, R. M. Jr. & Avise, J. C. (1992) Mitochondrial DNA phylogeographic differentiation among avian populations and the evolutionary significance of subspecies. *The Auk* **109:** 626-636.

Bandelt, H.-J., Forster, P. & Röhl, A. (1999) Median-Joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16:** 37-48.

Barrowclough, G. F. (1982) Geographic variation, predictiveness, and subspecies. *The Auk* **99:** 601-603.

Bense, U. (1995) Longhorn beetles. Illustrated key to the Cerambycidae and Vesperidae of Europe. Margraf Verlag, Weikersheim, 512 pp.

Bourtzis, K. & O'Neill, S. L. (1998) *Wolbachia* infections and arthropod reproduction. *Bioscience* **48:** 287-293.

Bruen, T. C., Philippe, H. & Bryant, D. (2006) A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172:** 2665-2681.

Cesari, M., Marescalchi, O., Francardi, V. & Mantovani, B. (2004) Taxonomy and phylogeny of European *Monochamus* species: first molecular and karyological data. *Journal of Zoological Systematics and Evolutionary Research* **43:** 1-7.

Cherepanov, A. I. (1983) Usachi Severnoj Azii (Lamiinae: Dorcadionini – Apomecynini). Izd. "Nauka", Sibirskoe Otdelenie, Novosibirsk. 223 pp. [In Russian]
Cracraft, J. (1983) Species concepts and speciation analysis. Pp. 159 – 187. Johnston, R.F. (ed.), Current Ornithology. Vol 1. Springer, Boston, MA.
Cranston, P. S. & Krosch, M. N. (2015) Nomenclatural corrections to Australian species of *Cricotopus* (Wulp) (Diptera; Chironomidae). *Zootaxa* **3980:** 298-300.
Danilevsky, M. L. (2012) Additions and corrections to the new Catalogue of Palaearctic Cerambycidae (Coleoptera) edited by I. Löbl and A. Smetana, 2010. Part. III. *Munis Entomology & Zoology* **7:** 109-173.
Danilevsky, M. L. (2017) Catalogue of Palaearctic Cerambycoidea. Updated 10.12.2017. Available at www.cerambycidae.net/catalog.pdf.
Dayrat, B. (2005) Towards integrative taxonomy. *Biological Journal of the Linnean Society* **85:** 407-415.
Dering, M. & Lewandowski, A. (2009) Finding the meeting zone: where have the northern and southern ranges of Norway spruce overlapped? *Forest Ecology and Management* **259:** 229-235.
Dominik, J. & Starzyk, J. R. (2004) Owady uszkadzające drewno. Wyd. PWRiL. 584 pp. [in Polish]
Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29:** 1969-1973.
Ehrlich, P. R. (1957) Systematists and subspecies. *The Lepidopterists' News* **11:** 155-157.
Engelstädter, J. & Hurst, G. D. (2009) The ecology and evolution of microbes that manipulate host reproduction. *Annual Review of Ecology and Systematics* **40:** 127-149.
Etxebeste, I., Sanchez-Husillos, E., Álvarez, G., Mas i Gisbert, H. & Pajares, J. (2015) Dispersal of *Monochamus galloprovincialis* (Col.: Cerambycidae) as recorded by mark-release-recapture using pheromone traps. *Journal of Applied Entomology* **140:** 485-499.
Evans, H. F., Moraal, L. G. & Pajares, J. A. (2004) Biology, ecology and economic importance of Buprestidae and Cerambycidae (pp. 447-474) Lieuter, F., Day, K. R., Battisti, A., Grégoire, J. C., Evans, H. F. (eds), Bark and Wood Boring Insects in Living Trees in Europe, a Synthesis. Kluwer, Dordrecht, The Netherlands.
Excoffier, L. & Lischer, H. E. L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10:** 564-567.
Farrell, B. D. (1998) "Inordinate fondness" explained: why are there so many beetles? *Science* **281:** 555-559.
Grenier, S., Barre, P. & Litrico, I. (2016) Phenotypic plasticity and selection: nonexclusive mechanisms of adaptation. *Scientifica* ID 7021701, 9 pages.
Grove, S. J. & Stork, N. E. (2000) An inordinate fondness for beetles. *Invertebrate Systematics* **14:** 733-739.
Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59:** 307-321.
Hall, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95-98.
Hambäck, P. A., Weingartner, E., Ericson, L., Fors, L., Cassel-Lundhagen, A., Stenberg, J. A. & Bergsten, J. (2013) Bayesian species delimitation reveals generalist and specialist parasitic wasps on *Galerucella* beetles (Chrysomelidae): sorting by herbivore or plant host. *BMC Evolutionary Biology* **13:** 92.
Haran, J., Rousselet, J., Tellez, D., Roques, A. & Roux, G. (2017) Phylogeography of *Monochamus galloprovincialis*, the European vector of the pinewood nematode. *Journal of Pest Science* **91**: 247-257.
Hellrigl, K. G. (1970) Die Bionomie der europäischen *Monochamus*-Arten (Coleopt., Cerambycid.) und ihre Bedeutung für die Forst- und Holzwirtschaft. *Redia 1970 - 71* **52:** 367-509.
Heyrovský, L. (1955) Tesařikoviti – Cerambycidae. Fauna ČSSR. Nakladatelství ČSAV, Praha, 5: 348 pp.
Hosokawa, T., Kikuchi, Y., Nikoh, N., Shimada, M. & Fukatsu, T. (2006) Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. *PloS Biology* **4:** e337.
Hu, S.-j., Ning, T., Fu, D.-y., Haack, R. A., Zhang, Z., Chen, D.-d., Ma, X.-y. & Ye, H. (2013) Dispersal of the Japanese pine sawyer, *Monochamus alternatus* (Coleoptera: Cerambycidae), in mainland China as inferred from molecular data and associations to indices of human activity. *PLoS One* **8:** e57568.
Hurst, G. D. D. & Jiggins, F. M. (2000) Male-killing bacteria in insects: mechanisms, incidence and implications. *Emerging Infectious Diseases* **6:** 329-336.
Huson, D. H. & Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23:** 254-267.
International Code of Zoological Nomenclature 2000. Available at http://iczn.org/code.

Isaev, A. S., Rozhkov, A. S. & Kiselev, V. V. (1988) Czernyj pikhtovyj usach *Monochamus urussovi* (Fisch.) [Fir sawyer beetle *Monochamus urussovi* (Fisch)]. Izd. Nauka, Nowosibirsk. 271 pp. [In Russian]

Katoh, K., Kuma, K., Toh, H. & Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33:** 511-518.

Kawai, M., Shoda-Kagaya, E., Maehara, T., Zhou, Z., Lian, C., Iwata, R., Yamane, A. & Hogetsu, T. (2006) Genetic structure of pine sawyer *Monochamus alternatus* (Coleoptera: Cerambycidae) populations in Northeast Asia: consequences of the spread of pine wilt disease. *Environmental Entomology* **35:** 569-579.

Kikuchi, Y. (2009) Endosymbiotic bacteria in insects: their diversity and culturability. *Microbes and Environments* **24:** 195-204.

Knowles, L. L. & Carstens, B. C. (2007) Delimiting species without monophyletic gene trees. *Systematic Biology* **56:** 887-895.

Kondo, K., Inouye, D. W., Segawa, M. & Seki, T. (1986) A demographic analysis of mortality caused by the pine nematode (*Bursaphelenchus xylophilus*) and pine sawyer beetles (*Monochamus alternatus*) in pine forests in the Seto Inland Sea-side, Japan. *Oekologia* **68:** 321-326.

Koutroumpa, F. A., Rougon, D., Bertheau, C., Lieutier, F. & Roux-Morabito, G. (2013) Evolutionary relationships within European *Monochamus* (Coleoptera: Cerambycidae) highlight the role of altitude in species delineation. *Biological Journal of the Linnean Society* **109:** 354-376.

Kubisz, D., Kajtoch, Ł., Mazur, M. A. & Rizun, V. (2012) Molecular barcoding for central-eastern European *Crioceris* leaf-beetles (Coleoptera: Chrysomelidae). *Central European Journal of Biology* **7:** 69-76.

Lacey, L. A. & Goettel, M. S. (1995) Current developments in microbial control of insect pests and prospects for the early 21st century. *Entomophaga* **40:** 3-27.

Laeché, A. D. & Fugita, M. K. (2010) Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society B* **277:** 3071-3077.

Latałowa, M. & van der Knaap, W. O. (2006) Late Quaternary expansion of Norway spruce *Picea abies* (L.). Karst in Europe according to pollen data. *Quaternary Science Reviews* **25:** 2780-2805.

Lecocq, T., Rasmont, P., Harpke, A. & Schweiger, O. (2015) Improving international trade regulation by considering intraspecific variation for invasion risk assessment of commercially traded species: the *Bombus terrestris* case. *Conservation Letters* **9:** 281-289.

Librado, P. & Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:** 1451-1452.

Linit, M. J., Kondo, E. & Smith, M. T. (1983) Insects associated with the pinewood nematode, *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae), in Missouri. *Environmental Entomology* **12:** 467-470.

Löbl, I. & Smetana, A. (eds) (2010) Catalogue of Palaearctic Coleoptera. Volume 6. Chrysomeloidea. Apollo Books, Stenstrup. 924 pp.

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

May, R. M. (1992) How many species inhabit the Earth? *Scientific American* (October issue) 18-24.

Mayr, E. (1982) Of what use are subspecies? *The Auk* **99:** 593-595.

Mazur, M., Holecová, M., Lachowska-Cierlik, D., Lis, A., Kubisz, D. & Kajtoch, Ł. (2016) Selective sweep of *Wolbachia* and parthenogenetic host genomes – the example of the weevil *Eusomus ovulum*. *Insect Molecular Biology* **25:** 701-711.

Mikšić, R. & Korpič, M. (1985) Cerambycidae of Yugoslavia, Part III. Academy of Sciences and Arts of Bosnia and Herzegovina, Sarajevo. 148 pp. [in Serbian]

Miller, D. R., Dodds, K. J., Eglitis, A., Fetting, C. J., Hofstetter, R. W., Langor, D. W., Mayfield III, A. E., Munson, A. S., Poland, T. M. & Raffa, K. F. (2013) Trap lure blend of pine volatiles and bark beetle pheromones for *Monochamus* spp. (Coleoptera: Cerambycidae) in pine forests of Canada and the United States. *Journal of Economic Entomology* **106:** 1684-1692.

Montagna, M., Chouaia, B., Sacchi, L., Porretta, D., Martin, E., Giorgi, A., Lozzia, G. & Epis, S. (2014) A new strain of *Wolbachia* in an Alpine population of the viviparous *Oreina cacaliae* (Coleoptera:Chrysomelidae). *Environmental Entomology* **43:** 913-922.

Montagna, M., Mereghetti, V., Lencioni, V. & Rossaro, B. (2016) Integrated Taxonomy and DNA barcoding of alpine midges (Diptera: Chironomidae). *PloS One* **11:** e0159124.

Montagna, M., Kubisz, D., Mazur, M. A., Magoga, G., Ścibior, R. & Kajtoch Ł. (2016b). Exploring species-

level taxonomy in the *Cryptocephalus flavipes* species complex (Coleoptera: Chrysomelidae). *Zoological Journal of the Linnean Society* **179**: 92-109.
Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B. & Worm, B. (2011) How many species are there on Earth and in the Ocean? *PloS Biology* **9:** e1001127.
Moritz, C. (1994) Defining "Evolutionarily Significant Units" for conservation. *TREE* **9:** 373-375.
Nakamine, H. & Takeda, M. (2008) Molecular phylogenetic relationships of flightless beetles belonging to the genus *Mesechthistatus* Breuning (Coleoptera: Cerambycidae) inferred from mitochondrial COI gene sequences. *Journal of Insect Science* **8:** 70-80.
Nei, M. (1972) Genetic distance between populations. *The American Naturalist* **106:** 283-292.
Normark, B. B., Jordal, B. H. & Farrell, B. D. (1999) Origin of a haplodiploid beetle lineage. *Proceedings of the Royal Society B, Biological Sciences, London* **266:** 2253-2259.
Nowakowska, J. A., Łukaszewicz, J., Borys, M., Tereba, A., Konecka, A., Zawadzka, A., Sułkowska, M. & Zajączkowski, P. (2017) Pochodzenie drzewostanów świerkowych (*Picea abies* L. Karst.) z Puszczy Białowieskiej na tle regionu RDLP Białystok na podstawie analiz mitochondrialnego DNA [Origin of the Norway spruce (*Picea abies* L. Karst.) stands from Białowieża Forest at the level of Białystok Regional Directorate of the State Forests based on mitochondrial DNA analysis]. *Sylwan* **161:** 40-51.
Ohbayashi, N. & Ogawa, J. (2009) Phylogenetic analysis of the lamiine Genus *Anoplophora* and its relatives (Coleoptera, Cerambycidae) based on the mitochondrial COI gene. *Special Bulletin of the Japanese Society of Coleopterology, Tokyo* **7:** 309-324.
Painter, J. N., Siitonen, J. & Hanski, I. (2007) Phylogeographical patterns and genetic diversity in three species of Eurasian boreal forest beetles. *Biological Journal of the Linnean Society* **91:** 267-279.
Phillimore, A. B. & Owens, I. P. F. (2006) Are subspecies useful in evolutionary and conservation biology? *Proceedings of the Royal Society of London, Series B* **273:** 1049-1053.
Plavil'shhikov, N. N. (1958) Zhuki-drovoseki, chast' 3. Podsemejstvo Lamiinae, ch. 1. Fauna SSSR. Zhestkokrylye. Moskva- Leningrad 23(1): 592 pp. [in Russian]
Presgraves, D. C. (2010) The molecular evolutionary basis of species formation. *Nature Reviews Genetics* **11:** 175-180.
Ronquist, F., Teslenko, M., Mark, van der P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A. & Huelsenbeck, J. P. (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61:** 539-542.
Rossa, R., Goczał, J. & Tofilski, A. (2016) Within- and between-species variation of wing venation in genus *Monochamus* (Coleoptera: Cerambycidae). *Journal of Insect Science* **16**: 1-7.
Ryder, O. A. (1986) Species conservation and systematics: the dilemma of the subspecies. *TREE* **1:** 9-10.
Sama, G. (2002) Atlas of the Cerambycidae of Europe and Mediterranean area. Volume 1. Northern, Western, Central and Eastern Europe. British Isles and Continental Europe from France (Excl. Corsica) to Scandinavia and Urals. – Kabourek, Zlín. 173 pp.
Sanz Muñoz, S. (2010) Analysis of nuclear markers in two species with highly divergent mtDNA lineages in Iceland: ITS in *Crangonyx islandicus* and EF alpha in *Apatania zonella*. University of Iceland Life and Environmental Sciences.
Schlick-Steiner, B. C., Steiner, F. M., Seifert, B., Stauffer, C., Christian, E. & Crozier, R. H. (2010) Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual Review of Entomology* **55:** 421-438.
Schmitt, T. & Varga, Z. (2012) Extra-Mediterranean refugia: The rule and not the exception? *Frontiers in Zoology* **9:** Art num 22.
Shoda-Kagaya, E., Kawai, M., Maehara, T., Iwata, R. & Yamane, A. (2008) Genetic structure of *Monochamus alternatus* in Japan (pp. 235 – 242) Mota, M. M., Vieira, P. (eds), Pine Wilt Disease: A Worldwide Threat to Forest Ecosystems. – Springer, Dordrecht.
Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994) Evolution weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America* **87:** 651-701.
Slatkin, M. (1993) Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* **47:** 264-279.
Sláma, M. (2006) Icones Insectorum Europae Centralis: Coleoptera, Cerambycidae. *Folia Heyrovskyana, Series B* **4:** 1-40.
StatSoft, Inc. (2011) STATISTICA (data analysis software system), version 10. Available at www.statsoft.com.

Stauffer, C., Lakatos, F. & Hewitt., G. M. (1999) Phylogeography and postglacial colonization routes of *Ips typographus* L. (Coleoptera, Scolytidae). *Molecular Ecology* **8:** 763-773.

Steinert, M., Hentschel, U. & Hacker, J. (2000) Symbiosis and pathogenesis: evolution of the microbe-host interaction. *Naturwissenschaften* **87:** 1-11.

Steward, J. R. & Lister, A. M. (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution* **16:** 608-613.

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, London* **277:** 661-671.

Stouthamer, R., Breeuwer, J. A. & Hurst, G. D. (1999) *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. *Annual Review of Microbiology* **53:** 71-102.

Sukumaran, J. & Knowles, L. L. (2017) Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences* **114:** 1607-1612.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.-G. & Cosson, J.-F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7:** 453-464.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28:** 2731-2739.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30:** 2725-2729.

Tollefsrud, M. M., Latałowa, M., van der Knaap, W. O., Brochmann, C. & Sperisen, C. (2015) Late Quaternary history of North Eurasian Norway spruce (*Picea abies*) and Siberian spruce (*Picea obovata*) inferred from macrofossils, pollen and cytoplasmic DNA variation. *Journal of Biogeography* **42:** 1431-1442.

Tomminen, J. & Leppänen, P. (1991) The male genitalia of *Monochamus sutor* (L.), *M. galloprovincialis* (Olivier) and *M. urussovi* (Fischer von Waldheim) (Coleoptera, Cerambycidae). *Entomologica Fennica* **2:** 49-51.

Valentini, A., Miquel, C., Nawaz, M. A., Bellemain, E., Coissac, E., Pompanon, F., Gielly, L., Cruaud, C., Nascetti, G., Wincker, P., Swenson, J. E. & Taberlet, P. (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Molecular Ecology Resources* **9:** 51-60.

Vicente, C. S., Nascimento, F. X., Espada, M., Barbosa, P., Hasegawa, K., Mota, M. & Oliveira, S. 2013. Characterization of bacterial communities associated with the pine sawyer beetle *Monochamus galloprovincialis*, the insect vector of the pinewood nematode *Bursaphelenchus xylophilus*. *FEMS Microbiology Letters* **347:** 130-139.

Wallin, H., Schroeder, M. & Kvamme, T. (2013) A review of the European species of *Monochamus* Dejean, 1821 (Coleoptera, Cerambycidae) – with a description of the genitalia characters. *Norwegian Journal of Entomology* **60:** 11-38.

Wang, Q. (ed.) (2017) Cerambycidae of the World: Biology and Pest Management. Boca Raton, Taylor et Francis Group. xiv + 627 pp.

White, J. A., Kelly, S. E., Perlman, S. J. & Hunter, M. S. (2009) Cytoplasmic incompatibility in the parasitic wasp *Encarsiainaron*: disentangling the roles of *Cardinium* and *Wolbachia* symbionts. *Heredity* **102:** 483-489.

White, J. A., Kelly, S. E., Cockburn, S. N., Perlman, S. J. & Hunter, M. S. (2011) Endosymbiont costs and benefits in a parasitoid infected with both *Wolbachia* and *Cardinium*. *Heredity* **106:** 585- 591.

Wilson, E. O. & Brown, W. L. Jr. (1953) The subspecies concept and its taxonomic application. *Systematic Zoology* **2:** 97-111.

Winkler, H., Christie, D. & Nurney, D. (1995) Woodpeckers: A Guide to the Woodpeckers, Piculets and Wrynecks of the World. Pica Prees. 406 pp.

Yang, Z. (2015) The BPP program for species tree estimation and species delimitation. *Current Zoology* **61:** 854-865.

Zabalou, S. & Riegler, M., Theodorakopoulou, M., Stauffer, C., Savakis, C. & Bourtzis, K. (2004) *Wolbachia*-induced cytoplasmic incompatibility as a means for insect pest population control. *Proceedings of the National Academy of Sciences of the USA* **101:** 15042-15045.

Zchori-Fein, E. & Perlman, S. J. (2004) Distribution of the bacterial symbiont *Cardinium* in arthropods. *Molecular Ecology* **13:** 2009-2016.

Zhang, J., Kapli, P., Pavlidis, P. & Stamatakis, A. (2013) A general species delimitation method with applications to phylogenetic placements bioinformatics. *Bioinformatics* **29:** 2869-2876.

2.3.4 Appendix

Supplementary information available at http://www.senckenberg.de/arthropod-systematics.

2.3.5 Personal contribution to the work

Molecular species delimitation analyses through BPP and writing of the part of manuscript related to these analyses.

2.4 Research article

Molecular species delimitation of the Asian chestnut gall wasp biocontrol agent released in Italy

Montagna, M., Gonella, E., Mereghetti, V., Magoga, G., Ferrari, E., Pontini, M., Ferracini & C., Alma, A. (2018) Molecular species delimitation of the Asian chestnut gall wasp biocontrol agent. *Insect Systematics and Evolution* **50:** 327-345. doi: 10.1163/1876312X-00002188

2.4.1 Summary

Molecular species recognition and identification, based on the mitochondrial *cox1* and on the nuclear ITS2, were performed on individuals of *Torymus sinensis* collected in Italy, on its close relative *T. beneficus* and on native torymids. The automatic-gap-discovery (ABGD) analyses correctly separate almost all morphospecies. On the basis of *cox1*, individuals of late-spring *T. beneficus* clustered with *T. sinensis,* and those identified as early-spring *T. beneficus* were recognized as a separate entity. Whereas, *T. beneficus* ecotypes clustered with *T. sinensis* on the basis of ITS2. Coalescent tree-based methods confirmed these results. The *cox1*-based recognition of early-spring *T. beneficus* as a separate phylospecies led us to conclude that this taxon deserves to be treated as a valid species, whereas individuals identified as late-spring *T. beneficus* might be considered as part of *T. sinensis*. Morphological identification and BLAST analyses confirmed that no *T. beneficus* was imported into Italy to control *Dryocosmus kuriphilus.*

2.4.2 Manuscript

Introduction

The Asian chestnut gall wasp (ACGW), *Dryocosmus kuriphilus* Yasumatsu, 1951 (Hymenoptera: Cynipidae), native of China, was considered to be a major pest of chestnut (*Castanea* spp.) by the mid-20th century. It established itself in several countries, being reported in Japan (1941), Korea (1958), the United States (1974), Nepal (1999), and Canada (2012). In Europe, it was first recorded in Italy in 2002, and is now reported in 19 European countries (Delalić, 2016; EPPO, 2016; Michaelakis, 2016; Radócz, 2016).

This gall wasp induces the formation of greenish-red galls, which develop at the time of budburst in the following early-spring, suppressing shoot elongation and causing a gradual decline in the vigor

of the host plants (EFSA, 2010). Severe nut production losses due to the development of galls were estimated to reach up to 85% in northern Italy (Bosio et al., 2013; Battisti et al., 2014).

In Japan, the indigenous parasitoid *Torymus beneficus* Yasumatsu & Kamijo, 1979 (Hymenoptera: Torymidae) was used as a biological control agent of the ACGW; however, it was not able to suppress chestnut gall wasp populations, thus requiring the introduction of the congeneric species *T. sinensis* Kamijo, 1982 from China (Yara, 2006). These two species, *T. sinensis* and *T. beneficus*, are difficult to separate on the basis of either their morphology or ecological traits. The ratio of the ovipositor sheath length to the lateral length of the thorax (O/T ratio), and the emergence period were used to distinguish empirically the two species (Ôtake, 1987; Moriya et al., 1992). However, the O/T ratio was shown not to be completely reliable (Yara, 2004) and applicable for female adults only, and the discrimination based on the emergence period may be hampered by slight differences observed between the two species and by the fact that two ecotypes of *T. beneficus* have been described, namely early-spring and late-spring types, each showing different emergence periods (Yara et al., 2000). Molecular markers were developed to distinguish the two species, such as isozyme and DNA markers, the latter targeting nuclear and mitochondrial DNA (Yara, 2004; Yara, 2006; Yara & Kunimi, 2009). Additional problems with species discrimination are based on hybridization that was observed between *T. sinensis* and both ecotypes of *T. beneficus* (Toda et al., 2000; Yara et al., 2000; Yara et al., 2010). These hybrids are known to have intermediate morphological features in respect to the parental species. Hybridization between males of *T. sinensis* and females of *T. beneficus* was hypothesized on the basis of mtDNA showing a closer relationship between late-spring *T. beneficus* and *T. sinensis* than to early-spring *T. beneficus* (Yara, 2004). Hence, definitive differentiation between *Torymus* species and hybrids has not yet been achieved (Yara & Kunimi, 2009; Yara, 2014). Different possible hypotheses could be formulated to explain the phylogenetic pattern reported in Figure 2 in Yara (2004), mainly: *i*) inadequate taxonomy, *viz T. beneficus* late spring ecotype is synonym of *T. sinensis*, as suggested by the great variability on the O/T ratio of these species (Yara, 2004); *ii*) incomplete lineage sorting; and *iii*) the occurrence of endosymbiont-mediated mtDNA introgression causing discordance between species and gene trees in the *T. sinensis – T. beneficus* group, as in the case of the parasitoid wasps of the genus *Diplazon* and the alpha-proteobacterium *Wolbachia* (Klopfstein et al., 2016).

After 15 years from the first report of the ACGW in Europe, and despite the species richness of the recruited parasitoid community, parasitism by native natural enemies is still low (Matošević & Melika, 2013; Quacchia et al., 2013; Alma et al., 2014; Francati et al., 2015; Colombari & Battisti, 2016). For this reason, a classical biological control approach was undertaken in Europe similar to those in Japan and in the USA (Moriya et al., 2003; Cooper & Rieske, 2011). *Torymus sinensis*

specimens imported from Japan were released in Italy starting in 2005 (Quacchia et al., 2008; Ferracini et al., 2015a), and afterwards in other European countries (Paparella et al., 2016). It proved to settle successfully in the chestnut-growing areas containing the ACGW outbreaks especially in northern Italy where it was initially released. Moreover, after a ten-year period, a differentiation in observed behaviors was recorded (e.g., prolonged diapause and host range expansion; Ferracini et al., 2015a; Ferracini et al., 2017).

Even if no evidence of divergent emergence periods for *T. sinensis* was ever recorded in Italy, but due to the difficulty in discriminating *T. beneficus*, *T. sinensis*, and hybrids, investigations on the biocontrol agent of the ACGW released in Italy were carried out. In particular, using molecular approaches and species delimitation tools, three major goals are addressed in the present study: *i*) test the congruence between morphospecies and phylospecies in the introduced and indigenous *Torymus* species; *ii*) establish whether *T. beneficus* individuals or hybrids were initially released in Italy together with Japan-imported *T. sinensis* specimens; *iii*) investigate the presence of *Wolbachia* in the examined species to provide preliminary information regarding the distribution of this bacterium within *T. sinensis* – *T. beneficus* late spring ecotype.

Materials and Methods

Ethics statement

All experiments were conducted in accordance with the legislation and guidelines of the European Union for the protection of animals used for scientific purposes (http://ec.europa.eu/environment/chemicals/lab_animals/legislation_en.htm).

All experimental protocols using animals were approved by the *ad-hoc* Committee of DISAFA of the University of Torino. In addition, all necessary permits were in hand when the research was conducted.

Insects sampling, manipulation and morphological identification

From 2006 to 2016, putative *T. sinensis* adults emerging from withered chestnut galls collected in Italy and Japan were used for this study, along with native species of *Torymus* emerging from withered oak and chestnut galls collected in Italy, as described in Table 1 and Figure 1. (Fig. 1 here) More precisely, regarding *T. sinensis*, the DNA was extracted from: *i*) specimens emerged in Italy from chestnut withered galls collected in Japan and provided by the National Agricultural Research Center in Tsukuba (Quacchia et al., 2008); *ii*) the progeny of the adults used in the first releases in Italy which occurred during 2006 and 2008; *iii*) individuals emerging from withered chestnut galls

collected in Piedmont (Italy) between 2014 and 2016, where the parasitoid was released for the first time, together with specimens emerging after a two-year diapause; and iv) three individuals, collected from oak galls of the gall wasp *Biorhiza pallida* (Olivier, 1791). Native torymids, emerging from galls of *D. kuriphilus*, *B. pallida*, *Cynips divisa* Hartig, 1840 and *Cynips quercusfolii* Linnaeus, 1758 collected in Aosta Valley, Liguria, and Piedmont regions of north-western Italy, were also included in the present study. Collections were performed both by hand and with the aid of lopping shears. Galls were isolated in plastic containers and stored in outdoor conditions until parasitoid emergence, according to the method described by Ferracini et al. (2015b).

Among the emerged parasitoids, torymid species were identified by using specific dichotomous keys reported in De Vere Graham & Gijswijt (1998) and Kamijo (1982), and by comparison with voucher specimens deposited at the DISAFA-Entomology laboratory, Italy. We were able to include *T. affinis* (Fonscolombe, 1832), *T. auratus* (Müller, 1764), *T. cyaneus* Walker, 1847, *T. flavip*es (Walker, 1833), and *T. geranii* (Walker, 1833) (Table 1).

Table 1. Samples analyzed in the study.

Species	♀♀	♂♂	*N**	Host gall wasp	Emergence	Gall provenience[a]	Vouchers/Accession no.	Marker	Reference
T. sinensis	5	6	11	*D. kuriphilus*	April 2006	IT - Piedmont Region	TsP06_1-11	*cox1*, ITS2	This study
"	5	3	8	"	May 2008	JP - Ibaraki Prefecture	TsJ08_1-8	"	"
"	3	3	6	"	May 2008	IT - Piedmont Region	TsP08_1-6	"	"
"	0	3	3	*B. pallida*	May 2013	"	TsPBp13_1-3	"	"
"	4	8	12	*D. kuriphilus*	March 2014	"	TsPd14_1-12	"	"
"	8	5	13	"	April 2014	"	TsP14_1-9, 11-14	"	"
"	7	4	11	"	April 2015	"	TsP15_1-11	"	"
"	7	8	15	"	April 2016	"	TsP16_1-15	"	"
T. affinis	2	0	2	"	March 2010	"	Taf10_1-2	"	"
"	3	0	3	*B. pallida*	March 2014	IT - Aosta Valley Region	Taf14_1-3	"	"
"	0	1	1	"	October 2014	"	TafA14_1	"	"
T. auratus	0	2	2	"	April 2014	IT - Liguria Region	Tau14_1-2	"	"
T. cyaneus	1	1	2	*D. kuriphilus*	May 2010	IT - Piedmont Region	Tc10_1-2	"	"
"	1	1	2	*C. divisa*	October 2014	"	Tc14_1-2	"	"
T. flavipes	0	3	3	*D. kuriphilus*	June 2014	IT - Aosta Valley Region	Tf14_1-3	"	"
T. geranii	1	0	1	"	August 2012	IT - Liguria Region	Tg12C_1	"	"
"	0	1	1	*C. quercusfolii*	April 2015	IT - Piedmont Region	Tg15Q_1	"	"
T. affinis	-	-	1	Oak cynipid	-	HU - Vas County	HM574341	*cox1*	Kaartinen et al. 2010
T. auratus	-	-	1	"	-	HU - Jász-Nagykun-Szolnok County	HM574340	"	"
T. beneficus late spring	-	-	4	*D. kuriphilus*	April 1988	JP - Ibaraki Prefecture	AB070476-AB070479	"	Yara 2004
T. beneficus early spring	-	-	4	"	March 1991	JP - Fukushima Prefecture	AB070473-AB070475, AB070480	"	"
"	-	-	4	"	March–April 1988	JP - Ibaraki Prefecture	AB070493 -AB070496	"	"
"	-	-	2	"	April 1993, 1997	JP - Nagano Prefecture	AB070497 - AB070498	"	"
"	-	-	1	"	March 1991	JP - Hiroshima Prefecture	AB070499	"	"
"	-	-	3	"	March–April 1993	JP - Kumamoto Prefecture	AB070501 - AB070503	"	"
"	-	-	1	"	March 1993	JP - Aomori Prefecture	AB070504	"	"
T. cyaneus	-	-	1	Oak cynipid	-	FI - Varsinais-Suomi Region	HM574245	"	Kaartinen et al. 2010
T. flavipes	-	-	1	-	-	SP	JQ416941	"	Stone et al. 2012
T. geranii	-	-	1	Oak cynipid	-	HU - Veszprém County	HM574309	"	Kaartinen et al. 2010
T. sinensis	-	-	4	*D. kuriphilus*	March–April 1993	CN - Hebei Province	AB070482 - AB070483, AB070485 - AB070486	"	Yara 2004
"	-	-	2	"	April 1993	CN - Liaoning Province	AB070484, AB070487	"	"
"	-	-	1	"	April 1992	KR - Kyongsangnam-do Province	AB070488	"	"
"	-	-	3	"	April 2000	KR - Kangwon-do Province	AB070489 - AB070491	"	"
"	-	-	1	"	April 1996	JP - Shimane Prefecture, Oki Islands	AB070492	"	"

Torymus sp.	-	-	1	*Celticecis japonica*	May 2000	JP - Chiba Prefecture	AB070500	"	"
"	-	-	1	*D. kuriphilus*	April 1993	JP - Aomori Prefecture	AB070481	"	"
T. flavipes	-	-	1	Oak cynipid		FI - Varsinais-Suomi Region	HM574237	ITS2	Kaartinen et al. 2010
T. geranii	1	0	1	*D. kuriphilus*	1997	JP - Ibaraki Prefecture	AB200280	"	Yara 2006
T. beneficus early spring	1	0	1	"	1988	JP - Ibaraki Prefecture	AB200271	"	"
T. beneficus late spring	1	0	1	"	1988	JP - Ibaraki Prefecture	AB200272	"	"
T. sinensis	-	-	1	"	1993	CN - Hebei Province	AB200274	"	"
Megastigmus dorsalis	-	-	1	-	-	CN - Guizhou	AY317240	*cox1*	Chen et al. 2004
"	-	-	1	Oak cynipid	-	-	GU123292	ITS2	Nicholls et al. 2010

[a]: country abbreviations according with Country codes ISO 3166.

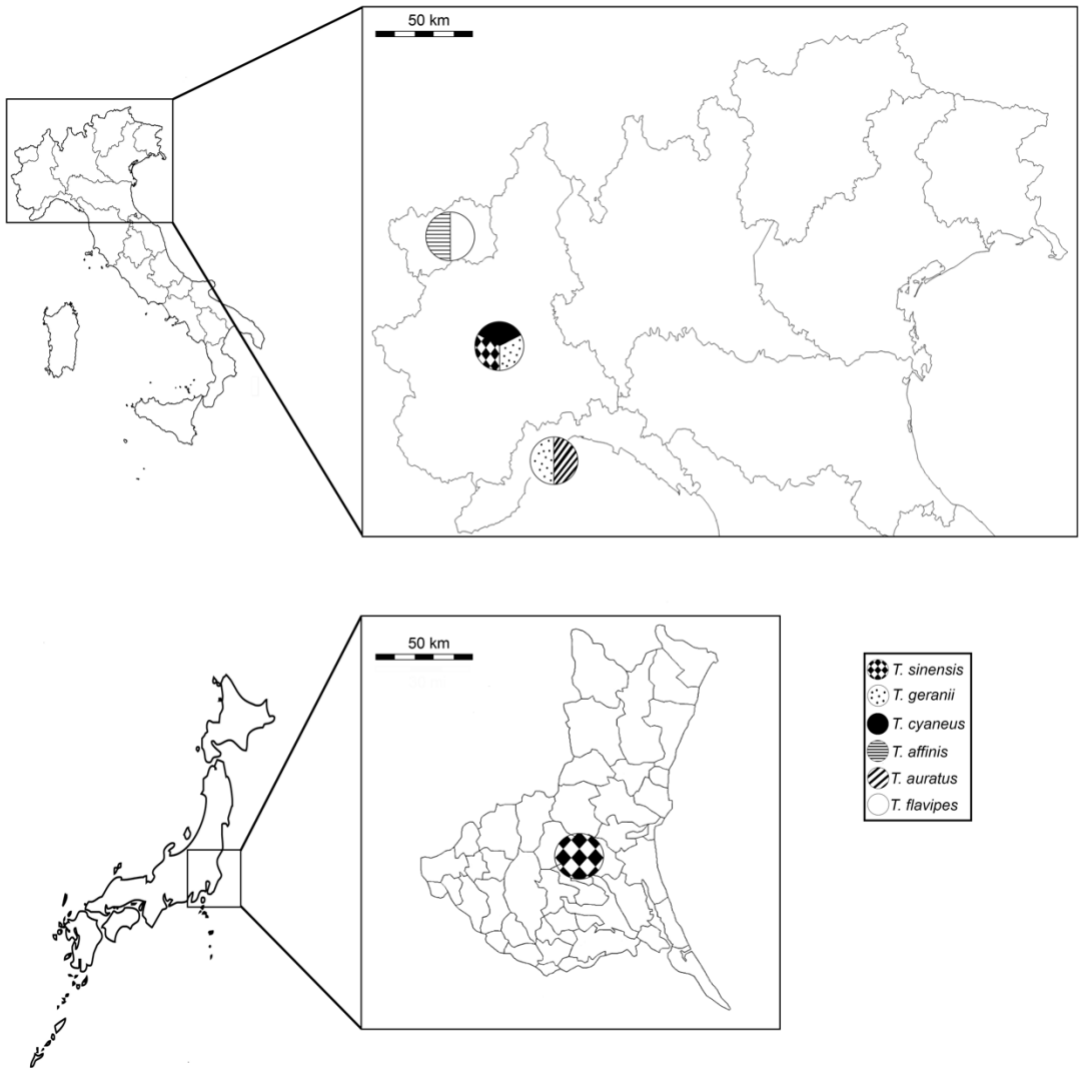

Figure 1. Collection sites of the specimens analyzed in the present study.

DNA extraction, PCR and sequencing

Chelex DNA extraction according to Kaartinen et al. (2010) was performed on metathoracic legs, dissected from the specimens. All the extracted specimens were stored in ethanol and preserved in the DISAFA collections with vouchers reported in Table 1. Samples were subject to amplification of the mitochondrial cytochrome c oxidase subunit 1 (*cox1*) gene, and the nuclear region of the internal transcribed spacer 2 (ITS2) between *5.8S* rRNA and *LSU* rRNA genes. For *cox1* amplification, primers COI pF2 (Kaartinen et al. 2010) and HCO (Folmer et al. 1994) were used with the following conditions: an initial denaturation step of 2 min at 94°C, followed by 5 cycles of 30 sec at 94°C, 1 min at 45°C, and 1 min at 72°C, then 35 cycles of 30 sec at 94°C, 1 min at 50°C, and 1 min at 72°C; a final extension step of 7 min at 72°C was applied. PCR for ITS2 was carried out with primers ITS2f/r as described by Campbell et al. (1993). The obtained PCR products were

purified with the commercially available kit (GenElute™ PCR Clean-Up Kit, Sigma-Aldrich) and sequenced (Genechron, Rome, Italy).

Electropherograms were manually checked using Geneious Pro 8.1 (Biomatters Ltd., Auckland, New Zealand); no heterozygotes were recovered in ITS2 electropherograms using a double peaks similarity of 95%. The nucleotide sequences were deposited in the European Nucleotide Archive with accession numbers LT821524 - LT821619 and LT821620 - LT821715, respectively for *cox1* and ITS2. To confirm the morphological identification of the specimens, *cox1* gene sequences were subjected to Nucleotide BLAST analysis against nr database (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

Orthologous sequences from *Torymus* spp., with the addition of *Bootanomyia* (=*Megastigmus*) *dorsalis* (Fabricius, 1798) as outgroup, were retrieved from GenBank (Table 1). The *cox1* dataset was aligned at the amino acid level using MUSCLE (Edgar, 2004) and back translated at the nucleotide level, whereas the ITS2 dataset was aligned using MAFFT (Katoh et al., 2005) with E-INS-i strategy. Poorly aligned positions in ITS2 alignment were trimmed with Gblocks (Castresana, 2000) using the strategy adopted by Sassera et al. (2010) for a less stringent selection.

Species delimitation analyses

Different methods, with no *a priori* information on the morphospecies, were adopted as described in previous studies (Montagna et al., 2016a, b). Briefly, the automatic barcode gap discovery tool (ABGD; Puillandre et al., 2012) and coalescent tree-based methods represented by the generalized mixed Yule-coalescent model (GMYC; Pons et al., 2006; Fujisawa & Barraclough, 2013) and its Bayesian implementation (bGMYC; Reid & Carstens, 2012) were used. These methods were extensively used to recognize and delimit species (e.g., Cranston & Krosh, 2015; Lecocq et al., 2015) as well as support the description of new insect taxa (e.g., Montagna et al., 2013; Montagna et al., 2016b).

ABGD analyses were performed on *cox1* and ITS2 aligned sequence datasets through the web interface (http://wwwabi.snv.jussieu.fr/public/abgd) with the uncorrected *p* distance (Collins et al., 2012; Srivathsan et al., 2012) and default settings for the remaining parameters.

The ultrametric tree required as input by the coalescent tree-based methods of species delimitation was inferred with BEAST (Drummond et al., 2012). Each alignment was analyzed with JModelTest 2 (Darriba et al., 2012) and the most suitable model of nucleotide evolution was selected using the Bayesian Information Criterion (Sullivan & Joyce, 2005). Identical haplotypes of conspecific specimens were removed from the datasets in order to outflank zero-length internal nodes in the trees. Phylogenetic reconstruction was achieved under the Bayesian framework using BEAST

(Drummond et al., 2012), three independent runs were performed for each dataset with the following settings: length of the Markov chain of 10^8 generations; trees and parameters sampled every 1,000 generations; models of nucleotide evolution as obtained by the model selection, *viz* GTR (Lanave et al., 1984) with gamma distribution (Γ) and proportion of invariable sites (I) in the case of *cox1*, K80 (Kimura, 1980) + I in the case of ITS2; uncorrelated relaxed clock type (Drummond et al., 2006); tree prior set on the Birth-Death Incomplete Sampling (Stadler, 2009); other prior parameters were set to default values. The convergence of the runs was visually examined by TRACER (Drummond et al., 2012), and the burn-in fraction estimated accordingly. After the removal of the burn-in trees fraction, trees were resampled at lower frequencies in order to obtain 5,000 trees for each run, and then resampled trees were merged and the maximum clade credibility tree was obtained using TreeAnnotator (Drummond et al., 2012).

Single-threshold GMYC analyses and bGMYC analyses were performed using the R packages SPLITS and bGMYC respectively. The maximum clade credibility ultrametric tree previously inferred was used as input for GMYC; whereas, in the case of bGMYC, 30 randomly sampled trees among the 15,000 merged BEAST trees were used as input. The parameters of bGMYC analysis were: Markov Chain Monte Carlo length of 100,000 generations, sampled every 100 generations and a burn-in of 800 trees, leading to the estimation of the statistics on a sample of 6,000 trees.

Pairwise nucleotide *p*-distance within and between taxa were calculated using MEGA 6 (Tamura et al., 2013), missing data and gaps were excluded in the pairwise distance estimation and the standard deviation was estimated by 500 bootstrap pseudo replicates.

Wolbachia detection

In order to provide preliminary information regarding the distribution of *Wolbachia* within the *T. sinensis – T. beneficus* group, DNA isolates from all *T. sinensis* specimens were tested for the presence of this bacterium by PCR targeting the 16S rRNA gene by using the W-Spec f/r primer pair as previously described (Werren & Windsor, 2000), after whole body DNA extraction according to Raddadi et al. (2011). None of the PCRs targeting *Wolbachia* genes led to positive amplifications, with the exception of the positive control, and thus this result will not be reported in the *Results and Discussion* section but only commented in the *Conclusion*.

Results

Datasets description and molecular species identification

DNA was extracted from a total of 96 adult specimens (47 females and 49 males) collected from different localities and attributed to the following six species: *T. affini*s (n = 6), *T. auratu*s (n = 2), *T. cyaneu*s (n = 4), *T. flavipe*s (n = 3), *T. geranii* (n = 2), and *T. sinensi*s (n = 79) (Table 1). A fragment of 427 bp of the mitochondrial *cox1* gene and ITS2 sequences, ranging from 478 bp of *T. sinensis* to 638 bp of *T. affinis*, were obtained for 95 individuals. The morphological identification of the specimens collected in Italy was confirmed by a BLAST search on *cox1* gene sequences (99%-100% of identity with conspecific specimens). According to previously published ITS2 profiles (Yara, 2006), all the *T. sinensis* collected in Italy possessed a homoallelic genotype, 57 individuals the 0/0, 14 individuals the -2/-2 and the specimen TsPBp13_1 collected in Piedmont from a gall of *B. pallida* the genotype -4/-4 (Table 2).

With the addition of orthologous gene sequences available in public repositories for the species *T. affinis*, *T. auratus*, *T. beneficus* (early- and late-spring ecotypes), *T. flavipes*, *T. geranii* and *T. sinensis*, the *cox1* and ITS2 datasets were composed of, respectively, 136 and 103 nucleotide sequences, which, resulted in 101 and 60 sequences for *cox1* and ITS2, respectively, after the removal of identical nucleotide sequences (as described in section *Species delimitation analyses* in Materials and Methods).

Table 2. Frequency of genotype, expressed as number of individuals, for ITS2 sequences of *T. sinensis*.

Species/Collecting locality	**Genotype**		
	-4/-4 +G((CT3,(AG3),(G)1)	-2/-2 ((CT)4,(AG)3)	0/0 ((CT)4,(AG)4)
T. sinensis (IT – Piedmont)	1	13	50
T. sinensis (JP – Ibaraki Prefecture)	0	1	7

Distance-based species delimitation

ABGD analysis performed on the aligned *cox1* gene sequences retrieved a perfect match between the initial and the recursive partitions at nucleotide divergence ranging from 0.3% to 1.3%, and nine groups were identified (or putative molecular species), highly congruent with morphological identification, (Fig. S1). Particularly, *T. affinis*, *T. auratus*, *T. cyaneus*, *T. flavipes*, *T. geranii*, *Torymus* sp. and *B. dorsalis* (outgroup) were recovered as separate entities; specimens identified in the literature as late-spring *T. beneficus* grouped with *T. sinensis*, while, all the individuals identified as early-spring *T. beneficus* belonged to a separate group. All the specimens collected in Italy (details are reported in section *Insects sampling, manipulation and morphological identification* in Materials and Methods) grouped with specimens identified as late-spring *T. beneficus* and as *T. sinensis* retrieved by previous studies. Increasing the prior intraspecific

nucleotide divergence to values between 2% and 3.6% (Figure S1), the initial and recursive partitions match in a second optimum, where the specimens identified as early-spring *T. beneficu*s were recognized in the same evolutionary unit with individuals of the late-spring ecotype and the closely related *T. sinensis*. The close relationship of *T. sinensis* and *T. beneficus* (both ecotypes) is also confirmed by the low between-taxa nucleotide *p*-distance of 4.8% (SD = 0.7%; Table 3). When the two ecotypes of *T. beneficus* are considered separately, the nucleotide *p*-distances between them and *T. sinensis* were 1.8% (SD = 0.4%) and 5.7% (SD = 0.9%), respectively in the case of late- and early-spring ecotypes (Table 3).

Similar results in the delimitation of the species were achieved also by ABGD analyses performed on the ITS2 dataset. Seven partitions were obtained from values of nucleotide *p*-distance ranging from 0.8% to 10%, with a match between the initial and recursive partitions (Fig. S2). In the case of ITS2, all the specimens identified as *T. sinensis* and the two specimens of *T. beneficus* (early- and late-spring) clustered within the same group (the achieved result is not attributable to the adopted trimming strategy since no nucleotides of these two taxa were removed from the alignment); a similar behavior was observed also for *T. auratus* and *T. geranii*. Interestingly, the specimen of *T. geranii* from Tsukuba (Japan, AB200280) was assigned to a separate group in respect to the Italian conspecific specimens. The latter result is in agreement with the extremely high value of within-taxa nucleotide *p*-distance recovered for *T. geranii* (average 21%, SD = 1.7%; Table 3), in the range of between taxa values. Except for *T. geranii*, the intraspecific nucleotide *p*-distance calculated on ITS2 spans from 2.4% (SD = 0.5%) to 0%, respectively in the case of *T. affinis* and *T. beneficus* (one individual for each of the two ecotypes) and *T. auratus* (Table 3); whereas the interspecific nucleotide *p*-distance range from 30.6% (SD = 2.1%) between *T. geranii* – *T. cyaneus* to 0.5% (SD = 0.2%) between *T. sinensis* and *T. beneficus* (Table 3).

Table 3. Within and between mean values of nucleotides *p*-distance[a].

***Torymus* species**	***p*-distance**								
	T. sinensis	*T. beneficus* LS	*T. beneficus* ES	*T. beneficus**	*T. auratus*	*T. geranii*	*T. affinis*	*T. cyaneus*	*T. flavipes*
T. sinensis	**1.3(0.3)/0.6(0.1)**	*1.8(0.4)*	*5.7(0.9)*	*4.8(0.7)*	*10.1(1.3)*	*9.9(1.3)*	*11.4(1.3)*	*12.3(1.5)*	*12.1(1.5)*
T. beneficus LS	0.5(0.2)	**0.9(0.2)/-**	*5.2(0.6)*		*10.5(1.2)*	*10.8(1.3)*	*11.9(1.3)*	*12.7(1.3)*	*11.8(1.3)*
T. beneficus ES	0.5(0.2)	0	**0.3(0.1)/-**		*11.2(1.3)*	*11.2(1.3)*	*11.7(1.3)*	*12.6(1.4)*	*12.1(1.3)*
*T. beneficus**	0.5(0.2)			**2.3(0.2)/0**	*11.1(1.2)*	*11.1(1.3)*	*11.8(1.2)*	*12.6(1.4)*	*12(1.3)*
T. auratus	24.2(2.2)	23.5(2.1)	23.3(2.1)	23.4(2.2)	**1.1(0.4)/0**	*9.7(1.3)*	*12.1(1.4)*	*13.1(1.5)*	*10.4(1.4)*
T. geranii	25.9(1.8)	25(1.8)	24.9(1.8)	25(1.8)	11.6(1)	**0.7(0.3)/21(1.7)**	*12(1.4)*	*12.2(1.4)*	*13(1.5)*
T. affinis	24.8(2)	24.4(2)	24.3(2)	24.3(2)	19.5(1.6)	19.5(1.6)	**1.6(0.4)/2.4(0.5)**	*11.5(1.4)*	*11.5(1.4)*
T. cyaneus	22.5(2.1)	21.9(2.1)	21.9(2.1)	21.9(2)	29.9(2.6)	30.6(2.1)	29(2.3)	**0.1(0.1)/0.1(0.1)**	*11.5(1.4)*
T. flavipes	20.4(2.1)	19.4(2)	19.6(2)	19.5(2)	8.5(1.4)	14.9(1.4)	15.2(1.8)	25.3(2.5)	**0.3(0.2)/1.2(0.4)**

[a]Number of base differences per site expressed as percentage. Above the diagonal are reported the mean values (italics) of between-taxa *p*-distance calculated on *cox1* gene; below the diagonal are reported the mean values of between-taxa *p*-distance calculated on ITS2; on the diagonal, in bold, are reported mean values of within-taxa *p*-distance calculated on *cox1* gene on ITS2, respectively on the right and on the left of /; standard deviation calculated on the base of 500 bootstrap is reported within brackets. *P*-distance values are calculated on nucleotide sequence databases without the removal of identical haplotypes.
[*]*p*-distance is calculated grouping sequences from *T. beneficus* early-spring (ES) and *T. beneficus* late-spring (LS).

Tree-based species delimitation

The topologies of the Bayesian consensus *cox1* and ITS2 trees, used as input for the tree-based species delimitation methods, were not totally congruent, as expected for closely related taxa (Figures 2 and 3). (Figures 2, 3 here)

On the basis of *cox1*, almost all the analyzed species were well supported and monophyletic, with a Bayesian posterior probability (BPP) of 1; the only exception is represented by the sequences identified as late-spring *T. beneficus* that clustered with *T. sinensis.* Two well supported clades (BPP of 0.97 and of 0.96), sub A and sub B in Figure 2, can be identified within the clade late-spring *T. beneficus* – *T. sinensis* (BPP = 1); however, a clear clustering pattern is not identifiable. All the specimens collected in Italy and identified as *T. sinensis* clustered within the clade *T. sinensis* – late-spring *T. beneficus*, it can thus be concluded that no specimens of early-spring *T. beneficus* were imported into Italy to control the ACGW through the withered chestnut galls. Even the specimen TsPBp13_1, with genotype -4+G corresponding to that of early-spring *T. beneficus*, clustered within the clade *T. sinensis* – late-spring *T. beneficus* on the basis of *cox1* sequence, as to be expected considering the results of BLAST analysis. Conversely, relationships among the species groups, especially those between the complex *T. sinensis* – *T. beneficus* and the remaining species are not resolved on the basis of *cox1* gene. Moreover, in the case of the ITS2 tree, the Bayesian posterior probability associated with morphospecies was high but *T. beneficus* and *T. geranii* resulted paraphyletic.

Species delimitation analyses performed by implementing coalescent tree-based approaches led to almost identical results for both *cox1* and ITS2 markers. Regarding the *cox1* dataset, the GMYC model exhibited a significantly better likelihood than the null model ($\log L_{GMYC} = 822.6$, $\log L_{NULL} = 805.8$, $2\Delta L = 33.5$, χ^2 test *p-value* < 0.0001). At the threshold between Yule and Coalescent models, *viz* cladogenesis – anagenesis, seven maximum likelihood clusters (95% CI [7, 8]) and a total of ten entities or putative molecular species (95% CI [10, 11]) were identified (Figure 2). Identical results were obtained by bGMYC, which identified 10 evolutionary units supported by a BPP ≥ 0.65 (Figure 2). These results are in accordance with those obtained by the distance-based approach. The morphospecies included in the present study, with the only exception of *T. sinensis* – *T. beneficus*, were identified by GMYC and bGMYC as belonging to separate entities. *Torymus beneficus*, as previously reported (Yara, 2004), resulted paraphyletic; the lineage with individuals assigned to early-spring phenotype resulted monophyletic and well isolated ($BPP_{bGMYC} = 0.75$), the sister of the clade *T. sinensis* – late-spring *T. beneficus* ($BPP_{bGMYC} = 0.65$). For the highest value of the confidence interval of entities recovered by GMYC (i.e., 11), obtained moving the Yule-Coalescent threshold towards the

tips, the clade *T. sinensis* – late-spring *T. beneficus* is split into two distinct clades (Figure 2) with a clustering pattern of individuals not explicable either in terms of their geographic origin nor in terms of ecology. All the specimens identified as late-spring *T. beneficus* are grouped in the clade sub A (BPP = 0.97; Figure 2), possibly corresponding to subgroup 3 in Figure 2 of Yara (2004). However, the fact that individuals of *T. sinensis* (AB07483, AB07484) belonging to subgroups 1 or 2 in Figure 2 of Yara (2004) clustered here within the clade sub A with late-spring *T. beneficus* does not support this possibility, but more likely, that *T. sinensis* – late-spring *T. beneficus* represents a single species.

On the basis of ITS2 seven maximum likelihood entities have been identified (Figure 3) ($logL_{GMYC}$ = 434.1, $logL_{NULL}$ = 431.1, $2\Delta L = 6$; *p-value* = 0.049), but a wider 95% confidence interval in respect to that achieved on the basis of the *cox1* marker is recovered (95% CI [3-34]). bGMYC reached almost identical results of those of GMYC, identifying five entities with BPP ≥ 0.65 and seven with BPP > 0.5 (Figure 3). It is noteworthy that both ecotypes of *T. beneficus* (represented by one specimen each) fall into a single group with *T. sinensis* (BPP = 0.67), confirming the results achieved by ABGD. Two separate and well-supported entities could be recognized within the clade composed of *T. sinensis* – *T. beneficus*. Apart from the *T. beneficus* – *T. sinensis* clade, *T. geranii* resulted paraphyletic, while *T. affinis*, *T. flavipes* and *T. cyaneus* resulted monophyletic and recognized as separate entities.

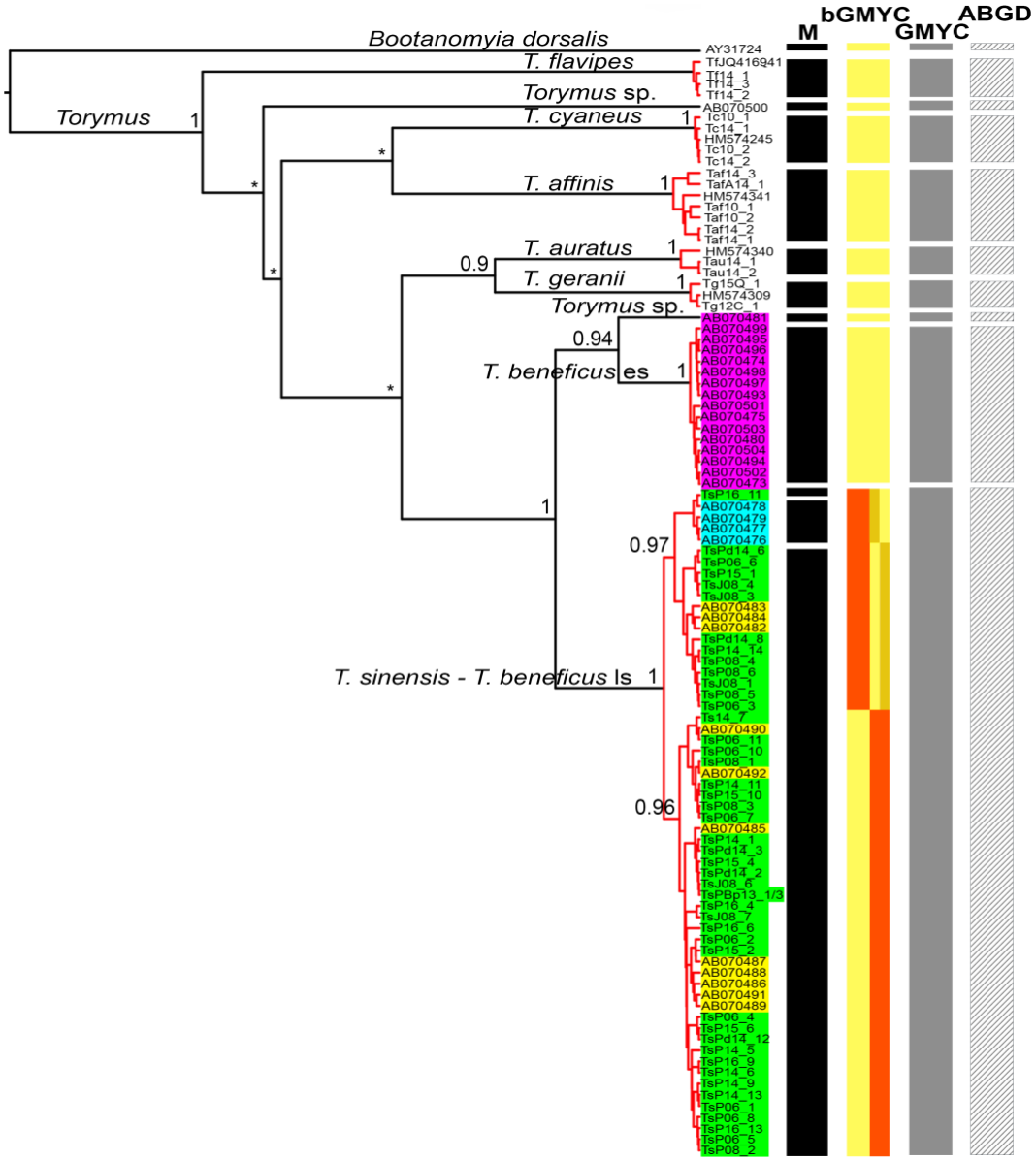

Figure 2. Species delimitation analysis based on cox1 gene sequences. A Bayesian ultrametric tree inferred from the cox1 gene sequence dataset, after the removal of identical haplotypes. *T. beneficus* – *T. sinensis* specimens are highlighted with different colors as follow: green = *T. sinensis*, from the present study; yellow = *T. sinensis*, from GenBank (Table 1); light blue = late-spring T. beneficus ecotype, from GenBank (Table 1); fuchsia = early-spring *T. beneficus* ecotype, from GenBank (Table 1). Clades corresponding to GMYC maximum likelihood clusters or putative molecular species are reported in red, and by dark grey vertical blocks; vertical black blocks indicate the identified morphospecies (M); putative molecular species identified by bGMYC are represented by vertical solid colored boxes, colors indicate support values of Bayesian posterior probability (BPP): 0.5 – 0.9 in orange, 0.9 – 0.95 in ochre, 0.95 – 0.99 in yellow; light grey texture boxes indicate putative molecular species identified by ABGD. On the main nodes of the phylogram are reported the values of BPP ≥ 0.7; * = bpp < 0.7.

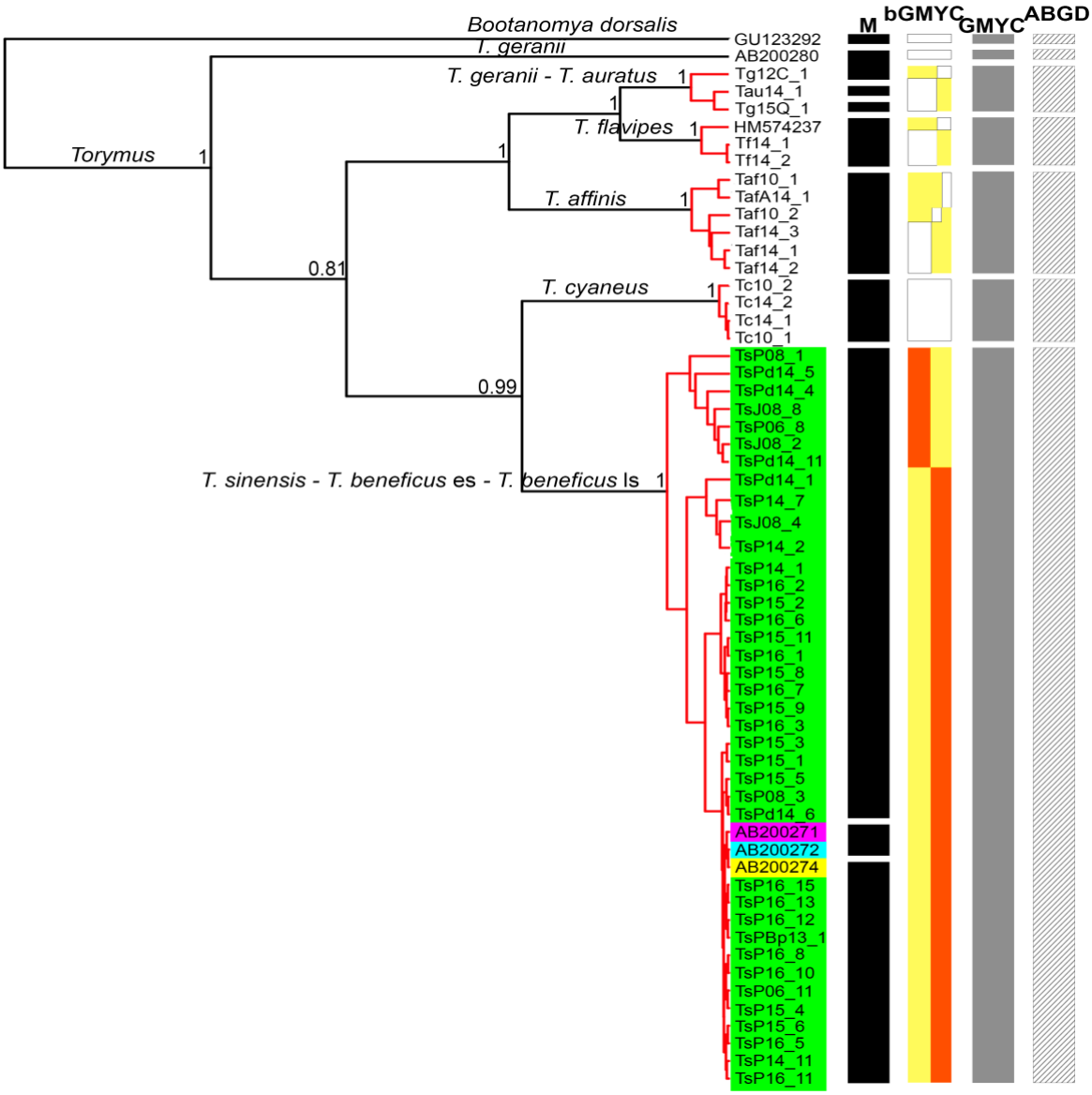

Figure 3. Species delimitation analysis based on ITS2 gene sequences. A Bayesian ultrametric tree inferred from the ITS2 gene sequence dataset, after the removal of identical haplotypes. *T. beneficus* – *T. sinensis* specimens are highlighted with different colors as follow: green = *T. sinensis*, from the present study; yellow = *T. sinensis*, from GenBank (Table 1); light blue = late-spring *T. beneficus* ecotype, from GenBank (Table 1); fuchsia = early-spring *T. beneficus* ecotype, from GenBank (Table 1). Clades corresponding to GMYC maximum likelihood clusters or putative molecular species are reported in red, and by dark grey vertical blocks; vertical black blocks indicate the identified morphospecies (M); putative molecular species identified by bGMYC are represented by vertical solid colored boxes, colors indicate support values of Bayesian posterior probability (BPP): 0.5 – 0.9 in orange, 0.95 – 0.99 in yellow, 0.99 – 1 in white; light grey texture boxes indicate putative molecular species identified by ABGD. On the main nodes of the phylogram are reported the values of BPP ≥ 0.7.

Discussion

Taking advantage of the recent introduction of *T. sinensis* in Italy as a biocontrol agent of the ACGW, the present study is mainly focused on testing the congruence between morphospecies and putative molecular species (or phylospecies) of the highly morphologically similar *T. sinensis – T. beneficus* complex as well as native torymids. The results achieved by the previous studies attempting to distinguish *T. sinensis* from *T. beneficus* using molecular markers (Yara, 2004; Yara, 2006; Yara & Kunimi, 2009) prompted us to adopt rigorous and recently developed species delimitation tools (i.e., ABGD, GMYC and bGMYC) in order to test the species status of these two taxa and that of the native torymids. Moreover, since the clustering pattern of *T. sinensis* and *T. beneficus* previously obtained by *cox1* gene sequences was congruent with endosymbiont-mediated mtDNA introgression, specimens of *T. sinensis* collected in Italy were tested for the presence of *Wolbachia*, widespread in invertebrates and recognized as the master manipulator of their reproduction (Werren et al., 2008). None of the analyzed individuals showed the presence of this bacterium. However, since topologies obtained using *cox1* and ITS2 are discordant in respect to *T. sinensis – T. beneficus* group, *viz T. sinensis* and *T. beneficus* (early- and late-spring) belong to the same clade using ITS2 but early-spring *T. beneficus* belong to a different clade using *cox1*, the presence of *Wolbachia* during the evolution of the group followed by a secondary loss cannot be excluded on the basis of the present study. Moreover, the presence of others vertically inherited endosymbionts, as "*Candidatus* Cardinium" (Zchori-Fein & Perlman, 2004), was not considered in this study.

Using *cox1* and ITS2 markers, all the adopted species delimitation approaches, represented by nucleotide distance and coalescent tree-based methods, agree in recognizing as separate phylospecies all the native morphospecies (Figures 2 and 3). Interestingly, the ITS2 sequence of *T. geranii* available in GenBank (AB200280) did not cluster with conspecific specimens collected in Italy, and due to its high value of nucleotide *p*-distance relative to the other conspecific (>20%), it could be considered as belonging to a different taxon (Figure 3). Regarding the *T. sinensis – T. beneficus* (early- and late-spring ecotypes) complex, the identification of all specimens of early-spring *T. beneficus* as a separate phylospecies by the adopted species delimitation methods on the basis of *cox1* marker let us hypothesize that this taxon deserves to be treated as valid species. Whereas, individuals identified as late-spring *T. beneficus* resulted in a well-delimited group together with all *T. sinensis* and might be considered as part of this latter species (Figure 2). This result is further confirmed by the low value of the *cox1* mean nucleotide *p-distance* between *T. sinensis* and the late-spring *T. beneficus* (1.8%, SD = 0.4), which is in the range of the intraspecific nucleotide distance values estimated in this study for the

other analyzed *Torymus* species (Table 3). On the basis of ITS2 marker *T. beneficus* (represented only by two individuals for which the corresponding *cox1* sequences are not available in public repositories) and *T. sinensis* (39 individuals) belong to the same evolutionary unit (Figure 3). The two well-supported entities identified by bGMYC within the *T. sinensis – T. beneficus* do not correspond to any apparent geographic or ecological pattern of clustering. The results achieved using this marker possibly suggests that ITS2 profile is not reliable to distinguish species and ecotypes within the *T. sinensis* species complex; however, since to our knowledge, *cox1* and ITS2 sequences of the same *T. beneficus* individual are not present in public repositories, a conclusion cannot be reached using the available data.

The values of nucleotide *p*-distance, the achieved tree topologies, and the species delimitation results are congruent with the following scenarios for the *T. sinensis – T. beneficus* complex: *i*) a recent origin of the taxa, in which a complete lineage sorting has not yet been achieved; *ii*) the presence of two delimited species, *viz.* early spring *T. beneficus* and late spring *T. beneficus – T. sinensis*, with the resulting incorrect identification of individuals assigned to *T. beneficus* late spring; or, *iii*) a single species requiring synonymy.

In conclusion, the achieved results are far from being conclusive and an integrative taxonomy analysis with an increased number of specimens and based on a multi-gene approach is required to definitively solve the boundary between these two sister species. Besides, with our study any individual of the early-spring *T. beneficus* ecotype was found to be imported from Japan to control the ACGW, *Dryocosmus kuriphilus*.

2.4.3 References

Alma, A., Ferracini, C., Sartor, C., Ferrari, E. & Botta, R. (2014) Il cinipide orientale del castagno: lotta biologica e sensibilità varietale. *Italus Hortus* **21**: 15-29.

Battisti, A., Benvegnù, I., Colombari, F. & Haack, R. A. (2014) Invasion by the chestnutgall wasp in Italy causes significant yield loss in *Castanea sativa* nut production. *Agricultural and Forest Entomology* **16**: 75-79.

Bosio, G., Armando, M. & Moriya, S. (2013) Verso il controllo biologico del cinipide del castagno. *Informatore Agrario* **4**: 60-64.

Campbell, B. C., Steffen-Campbell, J. D. & Werren, J. H. (1993) Phylogeny of the *Nasonia* species complex (Hymenoptera: Pteromalidae) inferred from an internal transcribed spacer (ITS2) and 28S rDNA sequences. *Insect Molecular Biology* **2**: 225-237.

Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**: 540-552.

Chen, Y., Xiao, H., Fu, J. & Huang, D. W. (2004) A molecular phylogeny of eurytomid wasps inferred from DNA sequence data of 28S, 18S, 16S, and COI genes. *Molecular Phylogenetics and Evolution* **31**: 300-307.

Collins, R. A., Boykin, L. M., Cruickshank, R. H. & Armstrong, K. F. (2012). Barcoding's next top model: an evaluation of nucleotide substitution models for specimen identification. *Methods Ecology and Evolution* **3**: 457-465.

Colombari, F. & Battisti, A. (2016) Native and introduced parasitoids in the biocontrol of *Dryocosmus kuriphilus* in Veneto (Italy). *EPPO Bulletin* **46**: 275-285.

Cooper, W. R. & Rieske, L. K. (2011) A native and an introduced parasitoid utilize an exotic gall-maker host. *Biological Control* **56**: 725-734.

Cranston, P. S. & Krosh, M. N. (2015) DNA sequences and austral taxa indicate generic synonymy of *Paratrichocladius* Santos-Abreu with *Cricotopus* Wulp (Diptera: Chironomidae). *Systematic Entomology* **40**: 719-732.

Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**: 772.

Delalić, Z. (2016) Prvi nalaz karantinske štetočine kestenove ose šiškarice (*Drycosmus kuriphilus*) u Bosni i Hercegovini. *Biljni lekar* **44**: 58-64.

Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**: e88.

Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**: 1969-1973.

De Vere Graham, M. W. R. & Gijswijt, M. J. (1998) Revision of the European species of *Torymus dalman* (s. Lat.) (Hymenoptera: Torymidae). *Zoologische verhandelingen/ uitgegeven door het Rijksmuseum van Natuurlijke Historie te Leiden (Ministerie van Cultuur, Recreatie en Maatschappelijk Werk*) **317**: 1-202.

Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.

EFSA Panel on Plant Health (PLH) (2010) Risk assessment of the oriental chestnutgall wasp, *Dryocosmus kuriphilus* for the EU territory on request from the European Commission. *EFSA Journal* **8**: 1619.

EPPO (2016) PQR – EPPO database on quarantine pests (available online) http://eppo.int

Ferracini, C., Gonella, E., Ferrari, E., Saladini, M. A., Picciau, L., Tota, F., Pontini, M. & Alma, A. (2015a) Novel insight in the life cycle of *Torymus sinensis*, biocontrol agent of the chestnut gall wasp. *BioControl* **60**: 169-177.

Ferracini, C., Ferrari, E., Saladini, M. A., Pontini, M., Corradetti, M. & Alma, A. (2015b) Non-target host risk assessment for the parasitoid *Torymus sinensis* in Italy. *BioControl* **60**: 583-594.

Ferracini, C., Ferrari, E., Pontini, Hernández Nova, L. K., Saladini, M. A. & Alma, A. (2017) Post- release evaluation of the non-target effects of *Torymus sinensis*, a biological control agent of *Dryocosmus kuriphilus* in Italy. *BioControl* **62**: 445-456.

Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**: 294-299.

Francati, S., Alma, A., Ferracini, C., Pollini, A. & Dindo, M. L. (2015) Indigenous parasitoids associated with *Dryocosmus kuriphilus* in a chestnut production area of Emilia Romagna (Italy). *Bullettin of Insectology* **68**: 127-134.

Fujisawa, T. & Barraclough, T. G. (2013) Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent (GMYC) approach: a revised method and evaluation on simulated datasets. *Systematic Biology* **62**: 707-724.

Kaartinen, R., Stone, G. N., Hearn, J., Lohse, K. & Roslin, T. (2010) Revealing secret liaisons: DNA barcoding changes our understanding of food webs. *Ecological Entomology* **35**: 623-638.

Kamijo, K. (1982) Two new species of *Torymus* (Hymenoptera, Torymidae) reared from *Dryocosmus kuriphilus* (Hymenoptera, Cynipidae) in China and Korea. *Kontyû* **50**: 505-510.

Katoh, K., Kuma, K., Toh, H. & Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33**: 511-518.

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.

Klopfstein, S., Kropf, C. & Baur, H. (2016) *Wolbachia* endosymbionts distort DNA barcoding in the parasitoid wasp genus *Diplazon* (Hymenoptera: Ichneumonidae). *Zoological Journal of the Linnean Society* **177**: 541-557.

Lanave, C., Preparata, G., Saccone, C. & Serio, G. (1984) A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution* **20**: 86-93.

Lecocq, T., Dellicour, S., Michez, D., Dehon, M., Dewulf, A., De Meulemeester, T., Brasero, N., Valterova, I., Rasplus, J. & Rasmont, P. (2015) Methods for species delimitation in bumblebees (Hymenoptera, Apidae, *Bombus*): towards an integrative approach. *Zoologica Scripta* **44**: 281-297.

Matošević, D. & Melika, G. (2013) Recruitment of native parasitoids to a new invasive host: first results of *Dryocosmus kuriphilus* parasitoid assemblage in Croatia. *Bullettin of Insectology* **66**: 231-238.

Michaelakis, A., Papachristos, D., Chytas, D.A., Antonopoulou, P. D., Milonas, P. G. & Avtzis, D. N. (2016) First record of *Dryocosmus kuriphilus* in Greece. *EPPO Bulletin* **46**: 290-294.

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**: 243-253.

Montagna, M., Mereghetti, V., Lencioni, V. & Rossaro, B. (2016a) Integrated Taxonomy and DNA Barcoding of Alpine Midges (Diptera: Chironomidae). *PLoS One* **11**: e0149673.

Montagna, M., Kubisz, D., Mazur, M. A., Magoga, G., Ścibior, R. & Kajtoch, Ł. (2016b) Exploring species-level taxonomy in the *Cryptocephalus flavipes* species complex (Coleoptera: Chrysomelidae). *Zoological Journal of the Linnean Society* **179**: 92-109.

Moriya, S., Inoue, K. & Mabuchi, M. (1992) Interspecific relationship between an introduced parasitoid, *Torymus sinensis* Kamijo, as a biological control agent of the chestnut gall wasp, *Dryocosmus kuriphilus* Yasumatsu, and endemic parasitoid, *T. beneficus* Yasumatsu et Kamijo. *Acta Phytopathologica et Entomologica Hungarica* **27**: 479-483.

Nicholls, J. A., Preuss, S., Hayward, A., Melika, G., Csóka, G., Nieves-Aldrey, J. L., Askew, R. R., Tavakoli, M., Schönrogge, K. & Stone, G. N. (2010) Concordant phylogeography and cryptic speciation in two Western Palaearctic oak gall parasitoid species complexes. *Molecular Ecology* **19**: 592-609.

Ôtake, A. (1987) Comparison of some morphological characters among two strains of *Torymus beneficus* Yasumatsu et Kamijo and *T. sinensis* Kamijo (Hymenoptera: Torymidae). *Applied Entomology and Zoology* **22**: 600-609.

Paparella, F., Ferracini, C., Portaluri, A., Manzo, A. & Alma, A. (2016) Biological control of the chestnut gall wasp with *T. sinensis*: A mathematical model. *Ecological Modelling* **338**: 17-36.

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D., Vogler, A. P. & Hedin, M. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* **55**: 595-609.

Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G. (2012) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* **21**: 1864-1877.

Quacchia, A., Moriya, S., Bosio, G., Scapin, G. & Alma, A. (2008) Rearing, release and settlement prospect in Italy of *Torymus sinensis*, the biological control agent of the chestnut gall wasp *Dryocosmus kuriphilus*. *Biological Control* **53**: 829-839.

Quacchia, A., Ferracini, C., Nicholls, J. A., Piazza, E., Saladini, M. A., Tota, F., Melika, G. & Alma, A. (2013) Chalcid parasitoid community associated with the invading pest *Dryocosmus kuriphilus* in north-western Italy. *Insect Conservation Diversity* **6**: 114-123.

Raddadi, N., Gonella, E., Camerota, C., Pizzinat, A., Tedeschi, R., Crotti, E., Mandrioli, M., Bianco, P. A., Daffonchio, D. & Alma, A. (2011) '*Candidatus* Liberibacter europaeus' sp. nov. that is associated with and transmitted by the psyllid *Cacopsylla pyri* apparently behaves as an endophyte rather than a pathogen. *Environmental Microbiology* **13**: 414-426.

Radócz, L., Szilágyi, A., Nagy, M., Kovács, G. & Melika, G. (2016) Asian sweet chestnut gallwasp, *Dryocosmus kuriphilus* (Hymenoptera, Cynipidae): first record for Romania. *North-Western Journal of Zoology* **12**: 201-204.

Reid, N. M. & Carstens, B. C. (2012) Phylogenetic estimation error can decrease the accuracy species delimitation: A Bayesian implementation of the General Mixed Yule Coalescent model. *BMC Evolutionary Biology* **12**: 196.

Sassera, D., Lo, N., Epis, S., D'Auria, G., Montagna, M., Comandatore, F., Horner, D., Peretó, J., Luciano, A. M., Franciosi, F., Ferri, E., Crotti, E., Bazzocchi, C., Daffonchio, D., Sacchi, L., Moya, A., Latorre, A. & Bandi, C. (2011) Phylogenomic evidence for the presence of a flagellum and *cbb*(3) oxidase in the free-living mitochondrial ancestor. *Molecular Biology and Evolution* **28**: 3285-3296.

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

Stadler, T. (2009) On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology* **261**: 58-66.

Stone, G. N., Lohse, K., Nicholls, J. A., Fuentes-Utrilla, P., Sinclair, F., Schönrogge, K., Csóka, G., Melika, G., Nieves-Aldrey, J.-L., Pujade-Villar, J., Tavakoli, M., Askew, R. R. & Hickerson, M. J. (2012) Reconstructing community assembly in time and space reveals enemy escape in a western palearctic insect community. *Current Biology* **22**: 532-537.

Sullivan, J. & Joyce, P. (2005) Model selection in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* **36**: 445-466.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30**: 2725-2729.

Toda, S., Miyazaki, M., Osakabe, M. H. & Komazaki, S. (2000) Occurrence and hybridization of two parasitoid wasps, *Torymus sinensis* Kamijo and *T. beneficus* Yasumatsu et Kamijo (Hymenoptera: Torymidae) in the Oki islands. *Applied Entomology* **35**: 151-154.

Werren, J. H. & Windsor, D. (2000) *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? *Proceedings of the Royal Society B-Biological Science* **267**: 1277-1285.

Werren, J. H., Baldo, L. & Clark, M. E. (2008) *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology* **6**: 741-751.

Yara, K. (2004) Relationship between the introduced and indigenous parasitoids *Torymus sinensis* and *T. beneficus* (Hymenoptera: Torymidae) as inferred from mt-DNA (COI) sequences. *Applied Entomology and Zoology* **39**: 427-433.

Yara, K. (2006) Identification of *Torymus sinensis* and *T. beneficus* (Hymenoptera: Torymidae), introduced and indigenous parasitoids of the chestnut gall wasp *Dryocosmus kuriphilus* (Hymenoptera: Cynipidae), using the ribosomal ITS2 region. *Biological Control* **36**: 15-21.

Yara, K. (2014) Interaction between *Torymus sinensis* (Hymenoptera: Torymidae) and *T. beneficus*, introduced and indigenous parasitoids of the chestnut gall wasp *Dryocosmus kuriphilus* (Hymenoptera: Cynipidae). *Japan Agricultural Research Quarterly* **48**: 35-40.

Yara, K. & Kunimi, Y. (2009) Discrimination of *Torymus sinensis* Kamijo (Hymenoptera: Torymidae) and *T. beneficus* Yasumatsu et Kamijo and their hybrids by allele-specific PCR. *Applied Entomology and Zoology* **44**: 275-280.

Yara, K., Yano, E., Sasawaki, T. & Shiga, M. (2000) Detection of hybrids between introduced *Torymus sinensis* and native *T. beneficus* (Hymenoptera: Torymidae) in central Japan, using malic enzyme. *Applied Entomology and Zoology* **35**: 201-206.

Yara, K., Sasawaki, T. & Kunimi, Y. (2010) Hybridization between introduced *Torymus sinensis* (Hymenoptera: Torymidae) and indigenous *T. beneficus* (late-spring strain), parasitoids of the Asian chestnut gall wasp *Dryocosmus kuriphilus* (Hymenoptera: Cynipidae). *Biological Control* **54**: 14-18.

Zchori-Fein, E. & Perlman, S. J. (2004). Distribution of the bacterial symbiont *Cardinium* in arthropods. *Molecular Ecology* **13**: 2009-2016.

2.4.4 Personal contribution to the work

Molecular species delimitation analyses and writing of the part of manuscript related to these analyses.

2.5 Research article

Factors affecting the efficiency of molecular species delimitation: evidences from the species rich beetle family of Chrysomelidae

Magoga, G., Fontaneto, D. & Montagna, M. Factors affecting the efficiency of molecular species delimitation: evidences from the species rich beetle family of Chrysomelidae. (Manuscript in preparation)

2.5.1 Summary

Molecular species delimitation is an approach to taxonomy that, in the contest of global diversity loss, can be very usefull for accelerating species discovery. In this study, the effect of some intrinsic factors to the analysed data on molecular species delimitation efficiency has been tested. Real data, i.e. a set of more then 7,000 *cox1* sequences belonging to 250 leaf beetles species (Coleoptea: Chrysomelidae), were analysed through four species delimitation methods (3% nucleotide distance threshold, ABGD, GMYC, mPTP) in order to evaluate the possible effect of the following factors on the delimitation efficiency: i) the mean number of haplotypes per dataset; ii) the geographic distance among conspecific collection localities; iii) the difficulty of species morphological identification; iv) the taxonomic rank; v) the molecular species delimitation method adopted; vi) the number of morphospecies per dataset. From our results we observed a significant positive relation between methods efficiency and intraspecific geographic distance, observed for those datasets for which the intraspecific nucleotide distance did not increase proportionaly to the geographic distance. In addition, we observed that the higher is the number of species that are very difficult to be morphologically identified in a dataset, the lower is the methods efficiency, confirming that molecular species delimitation methods are able to underline morphological identification errors. Finally, a highly significant impact of the used method on delimitation efficiency was observed, with coalescent-tree based methods related to a lower identification efficiency, expecially on poor datasets in term of species and haplotypes per species.

2.5.2 Manuscript

Introduction

In the context of a rapid loss of global biodiversity, 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 et al., 2008; Monaghen et al., 2009; Mutanen et al., 2013). The single-locus DNA-based identification methods represented by DNA-barcoding and DNA-metabarcoding stand out as approaches adopted for biodiversity surveys (Telfer, 2015; deWaard et al., 2019). These methods, using as marker a standardized gene region, allow to speeding up organisms identification and in some cases overcome morphological taxonomy limits, for example making possible to detect organisms presence from their DNA spread in the environment (Thomsen & Willerslev, 2015; Ruppert et al., 2019). Beside molecular identification, biodiversity studies have been improved thanks to molecular species delimitation tools, which allow infering hypothetic species and/or evolutionary significant units from molecular data. Species delimitation molecular methods are used for both biodiversity investigation without a priori hypothesis on the possible species (Dinca et al., 2015; Gómez-Zurita, 2016) or, more frequently, as support to resolve taxonomic issue when other delimitation approaches are not effective (e.g., Montagna et al., 2016; Garcia-Melo, 2019; Plewa et al., 2018), thus contributing in the so-called "integrative taxonomy" approach (Dayrat, 2005). In the last decades many molecular species delimitation methods were proposed, among early methods we can find those analysing 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 known to be more accurate when taking information from more than one marker (Rubinoff & Holland, 2005; Dupuis et al., 2012), nevertheless, single-locus species delimitation is a widely adopted procedure (Pentinsaari et al., 2017), since quicker and cheaper than multi-locus data; in addition, the flourishing of DNA-barcoding and DNA-metabarcoding deeply supported the implementation of species delimitation methods on the barcoding data itself. Beside bacteria and fungi, the target organisms of biodiversity investigation through molecular taxonomy are frequently invertebrates, due to their challenging morphological taxonomy, to the presence of neglected groups that are historically poorly investigated (Ceccarelli et al., 2011), and/or to the species richness of some groups that often hide still undescribed taxa. For these organisms, the standard marker used is a segment of Cytochrome Oxidase subunit I (*cox1*) of ~650 bp, considered robust and reliable due to the region variability that allows researchers to discriminate organisms at species level and in particularly

to the highly consolidate primers for its amplification that are suitable for the majority of the groups (Hebert et al., 2003; Mioduchowska et al., 2018). One of the first methods proposed for delimiting species starting from *COX1* barcodes is a phenetic approach that relies on sequences pairwise nucleotide distances, which aims to find a value of nucleotide divergence that can be used as threshold to discriminate between intraspecific and interspecific nucleotide variability (Hebert et al., 2003; Meyer and Paulay, 2005). The delimitation using a distance threshold is accurate when a gap, the so called "barcoding gap", in the frequency distribution of intra- and inter-specific nucleotide distances is present. Since the barcoding gap is not always present, the threshold use is can be related to identification error, that increases even more intraspecific and interspecific distances overlap (Meyer & Paulay, 2005). The reported drawback could happen when a fixed threshold value is used to delimit groups of organisms with different evolutionary histories for which a proper threshold should be instead applied (Meyer & Paulay, 2005; Magoga et al., 2018). Starting from barcoding gap concept, the Automatic Barcode Gap Discovery method was proposed (e.g. ABGD, Puillandre et al., 2012). ABGD finds the distance at which the gap is located, also when intra-interspecific distribution overlap; moreover, it delimits taking in account the possibility that multiple thresholds are present among taxa. Some authors have criticized the use of distance thresholds as species delimitation method, since they do not account for evolutionary processes, (Hickerson et al., 2006). Alternative methods, as well as applied to DNA barcode datasets, are phylogenetic-coalescent based methods; among them, the most used are the Generalized Mixed Yule Coalescent method (GMYC, Pons et al., 2006; Fontaneto et al., 2007; Fujisawa and Barraclough, 2013), its Bayesian implementation bGMYC (Reid & Carstens, 2012), the Poisson Tree Process (PTP, Zhang et al., 2013) and the multirate PTP (mPTP, Kapli et al., 2017). All these methods require as input a phylogeny of taxa estimated from DNA sequences: briefly, GMYC looks for the maximum likelihood solution to separate the branches of a ultrametric tree between species (i.e. speciation), modelled by a Yule process (Yule, 1925), and within species processes based on neutral coalescent (Hudson, 1990); while PTP, basing on branch length, finds the transition point between intraspecific and interspecific process assuming a two parameter model that account for speciation and for the coalescent process.

Despite the fact that DNA-based species delimitation methods are widely used and consolidated, how they can be influenced by factors intrinsic to the analysed data need to be empirically evaluated. Many studies were carried out for testing which factors influence the results of species delimitation methods, but the majority of them were performed on simulated data where dataset characteristics were a priori defined in order to test a specific hypothesis. Among them, some evaluating 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 et al., 2012, Fujisawa & Barraclough, 2013; Dellicour & Flot, 2018). However, when analysing real data, almost no one of the previous factors can be accounted. Testing the factors affecting delimitation on real data can be helpful to define practical guiding lines to establish the correct study experimental design and/or to account for biases when reading delimitation results. Excluding Talavera et al. (2013), who analysed in detail the GMYC performances on 1,303 *cox1* sequences belonging to 172 species of Romanian butterfly, only an overview of factors affecting delimitation was provided from evaluations on empirical cases (Pentinsaari et al., 2017).

In light of this, in this study we investigated the influence of some biological and non-biological factors, such as the sampled number of haplotypes per species, the geographic distance between conspecific collection localities, the difficulty level of the morphological identification of the species, the dataset taxonomic rank, the molecular species delimitation method adopted and the number of morphospecies per dataset, on the results of species delimitation analyses using as model a collection of ~7,000 *cox1* sequences belonging to ~550 Euro-Mediterranean leaf beetles species (Coleoptea: Chrysomelidae). Chrysomelidae is one of the most specious families of Coleoptera, more than 45,000 species are present worldwide and 55,000 to 60,000 estimate to exist (Jolivet, 2015). The species within the family are almost all phytophagous at the adult stage, but very different in term of ecology such as in the trophic specialization (monophagy, oligophagy, poliphagy), habitat and width of distribution. In fact, in the Euro-Mediterranean area there are common species, widely distributed within the area (Canty et al., 2016), some cold adapted species living only in high altitude mountains (Brunetti et al., 2019) and some endemites (Biondi et al., 2013; Montagna et al., 2013). The set of sequences analysed in this work include almost ¼ of Euro-Mediterranean Chrysomelidae species, morphologically identified by specialists and representative in term of ecology of the taxa present within the family; for this reason, it can be considered suitable for testing factors influencing species delimitation efficiency. Despite a high number of molecular species methods is available, in this work we choosed to test some of the most used molecular species delimitation methods on single locus data (i.e. nucleotide distance threshold, ABGD, GMYC and mPTP).

Material and Methods

Datasets

In the present work the collection of *cox1* sequences developed and analysed in Magoga et al. (2018) was used as reference for the performed analyses. This collection is a selection of *cox1* sequences consisting of 7,237 high-quality and taxonomically assigned to species level sequences of 652 bp on average [range: 460–658] belonging to 542 species of Chrysomelidae (Insecta: Coleoptera), which could be considered a representative subsample of the Euro-Mediterranean Chrysomelidae, it includes approximately 24% of the species present in the area.

The Magoga et al. (2018) sequences set was divided into datasets according to taxa membership to the following main taxonomic levels: *i.* family, *ii.* subfamily, *iii.* genus (Fig. 1a).

All the datasets were aligned at codon level using MUSCLE in MEGA X (Kumar et al., 2018). The obtained alignments were used as input for the species delimitation analyses, *per se* or after the inference of a phylogenetic tree. The datasets used to infer phylogenetic trees, input of tree-based species delimitation methods, were collapsed in order to retain for each species only unique haplotypes with R software (R Core Team, 2019).

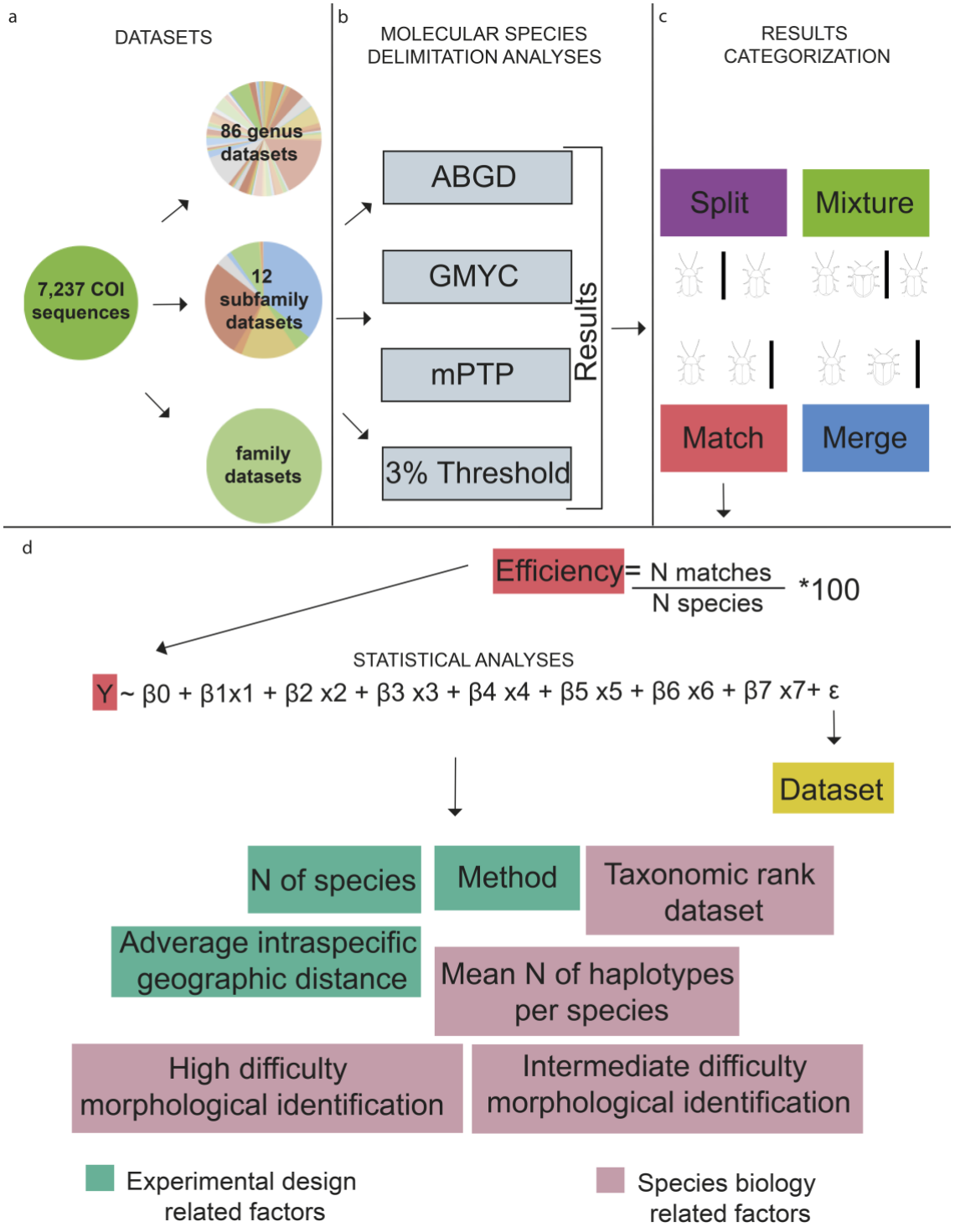

Figure 1. Experimental design of the study. The experimental part of this work can be summarized in the following steps: Datasets development (a); Molecular species delimitation analysis (b); Molecular species delimitation results categorization (c); Mixed Linear Model statistical analysis (d).

Molecular species delimitation analyses

In the present study, the most commonly used molecular species delimitation methods were adopted, two nucleotide distance-based and two coalescent tree-based molecular species delimitation methods (Fig. 1b): i. ABGD (Puillandre et al., 2012); ii. the 3% nucleotide distance threshold proposed by Herbert et al. 2003 iii. GMYC (Pons et al., 2006; Fontaneto et al., 2007; Fujisawa and Barraclough, 2013); iv. the Multi-rate Poisson tree processes (mPTP; Kapli et al., 2017).

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

The K2P pairwise nucleotide distance matrices, required for the species delimitation through 3% nucleotide distance threshold method, were estimated for each dataset using the R software library ape (Paradis et al., 2004). The R function *tclust* of package SPIDER was used to cluster nucleotide sequences at a distance threshold of 3%.

In order to perform GMYC and mPTP species delimitation analyses, an ultrametric tree was inferred using the software BEAST v 1.8 (Drummond et al., 2012) from each haplotype-reduced dataset including one orthologous sequence from the appropriate outgroup. For each dataset two to five independent MCMC runs were performed using the following parameters: Markov chain length from 30 Mgens to 300 Mgens depending on the runs convergence assessed examining the estimated sample size of each parameter of the model and a visual inspection of the likelihood by TRACER (Drummond et al., 2012); sampling of trees and parameters every 1,000-5,000 generations; models of nucleotide evolution as selected according to the Bayesian information criterion after the analysis performed by jModelTest 2 (Darriba et al., 2012); Yule process as speciation model (Yule, 1925), while other priors were 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 R function *gmyc* of the library SPLITS (Ezard et at., 2009). The method tests the hypothesis that the analysed samples belong to *n* independently evolving species; the comparison between the likelihood of the GMYC model with that of the null model (all samples belong the same independently evolving species) is

obtained through a log-likelihood ratio test. The delimitation hypothesis of GMYC is accepted when the likelihood of the GMYC model results significantly better than that of the null model.

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 M 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 obtained for each analysis were classified in the following categories adopting an ad-hoc developed R script (GM): 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. 1c).

Test of factors affecting species delimitation

The efficiency (*E*) of the delimitation was calculated as the number of match (*n*) obtained from each analysis on the overall number of morphospecies present in the dataset (*m*) as

$E = n/m\ 100$

The efficiency was used as dependent variable in a linear mixed model analysis for testing the effect of some species biology and study experimental design related factors on the delimitation results (Fig. 1d). The factors taken in account, independent variables of the model, were: i) the mean number of haplotypes per dataset (*mnhd*); ii) the median value of the geographic distance between conspecific collection localities (*mgdc*); iii) the difficulty level of the morphological identification of the species (*dl*); iv) the dataset taxonomic rank (*dtr*); v) the molecular species delimitation method adopted (*msdm*) vi) the number of morphospecies per dataset (*m*). Hereafter, each variable is described in detail.

Species biology related factors:

Mnhd: assuming that a greater number of haplotypes per species should better represent the overall species nucleotide variability (Goodall-Copestake et al., 2012), we can hypothesize that a more accurate delimitation is achieved when a high number of haplotypes per species is available. A higher number of haplotypes per species could better represent the coalescent process improving coalescent-

tree based methods performances, from the other hand, could lead distance based methods to merge two or more species in a group since it is more probable to sample intermediate haplotypes between species that hide the barcoding gap. The number of haplotypes for each species in a dataset was calculated and then the *mnhd* estimated.

Dtr: Even considering that taxonomy does not always follows systematics, taxonomic rank could be considered as a proxy of the phylogenetic relatedness of the species. In this work, we tested if molecular species delimitation methods results are influenced by the composition of the dataset in term of taxa analysing the sequences grouped in datasets basing on the taxonomic level they belong: family level dataset, subfamily level datasets and genus level datasets.

*Dl***:** Organisms morphological identification could be a difficult task, especially for some taxa, insects included, where the diagnostic characters for the identification of closely related species sometimes present only subtle differences (Magoga et al., 2018); in these cases, even taxonomists specialist of the subject group, could fail in the identification. Basically, when an organism is difficult to be morphologically identified, its assignment to a morphological species is more prone to error, in this case the estimate of species delimitation efficiency could be biased. In order to test this assumption, for each species considered in this study a level of difficulty in morphological identification was assigned. Level I: species easy to be identified morphologically since they present clear diagnostic morphological characters; level II: species that can be distinguished from morphologically similar species on the basis of subtle external morphological characters or on the basis of strong differences in the shape of genitalia (spermatheca or the median lobe of the aedeagus); level III: species that are completely identical to others in the external morphological characters and that can be distinguished from such similar species basing on subtle differences in the shape of genitalia. The level of difficulty was assigned to species basing on authors taxonomic competence and using specialized literature reporting Chrysomelidae species descriptions and dichotomous keys (Burlini, 1955; Müller, 1953; Doguet, 1994; Warchałowski, 2003). For each dataset, the percentage of species assigned to each difficulty level was calculated and registered in three variables, *dl I, dl II* and *dl III*.

Study experimental design related factors

Mgdc: isolation by distance (IBD) is a phenomenon that occurs within species when populations have a more or less continuous distribution and individuals a relatively low dispersal distance that lead the populations at opposite ends of the distribution to be more different compared to the others because of the longer absence of gene flow (Wright, 1978). The presence/absence of IBD is related to the

evolutionary history of each group that is not supposed to be known for performing a molecular species delimitation but may influence the delimitation results. For this reason, assuming the possible presence of IBD, that is a common phenomenon in phytophagous insects (Peterson & Denno, 1998), and consequently that increasing intraspecific geographic distance, intraspecific nucleotide distance increases, *mgdc* was included as factors possibly affecting species delimitation. For each species the median geographic distance among the collection localities of individuals was calculated from a pairwise geographic distance matrix using the R library geosphere (Hijmans, 2017); for each dataset the median of these values was calculated (*mgdc*).

*Msdm***:** Considering that different delimitation methods could give different delimitation results (Carstens et al., 2013), the influence of method choice on delimitation efficiency was tested. Two nucleotide distance-based (ABGD, 3% nucleotide distance threshold) and two coalescent tree-based methods (GMYC, mPTP) were used to analyse the developed datasets.

M: Number of morphological species included in each analysed dataset.

The presence of correlation between the independent variables of the model was tested through the R library PSYCH (Revelle, 2018). Since *dl I* and *dl II* were found to be strongly negatively correlated (-0.95 r), the former variable was excluded from the following analysis. The linear mixed model analysis for testing the influence of factors (*mnhd*, *mgdc*, *dl II*, *dl III*, *dtr*, *msdm*, *m*) on species delimitation efficiency was fitted using LMERTEST library (Kuznetsova et al., 2017); since measures were repeated four times on the same dataset, the dataset was defined as random-effect term. For assessing the significance of the independent variable a deviance analysis (Type II Wald chisquare tests) was fitted through R CAR package (Fox & Weisberg, 2019).

In addition, it has been tested if the results previously achieved are dependent on the size of the analysed dataset (i.e. the number of sequences). The previous linear mixed model was fitted on different subsets of datasets obtained retaining those that include a defined number of t_i sequences. The following size thresholds were applied: $t_1 > 5$ sequences; $t_2 > 10$ sequences; $t_3 > 20$ sequences; $t_4 > 40$ sequences; $t_5 > 80$ sequences; $t_6 > 160$ sequences.

Results

Datasets obtained and analyzed

The sequence set used and analysed in Magoga et al., 2018 was split in 98 datasets according to the taxonomic levels, resulting in one family, 12 subfamily and 86 genus datasets. Three genera datasets

were excluded from delimitation analyses since composed by only one *cox1* sequence. The family dataset is composed by 7,237 sequences, while subfamilies datasets are composed on average by 603 sequences (range: 3-2,690) and genera datasets by 87.2 sequences (range: 2-1,014). After the haplotypes reduction, the datasets were composed by 4,066, on average 338.8 (range: 3-1,456) and on average 50.1 (range 2-584) sequences in family, subfamilies and genera datasets, respectively. In median 6 sequences per species are present, but some species are more represented than others, e.g. *Gonioctena olivacea* (238 sequences, 118 haplotypes), *Longitarsus ordinatus* (164 sequences, 51 haploypes, *Calomicrus circumfusus* (161 sequences, 60 haplotypes).

GMYC and mPTP analyses were performed on 94 datasets (one family, 12 subfamilies and 81 genera datasets), two genera datasets were excluded since composed of only one sequence after haplotypes collapsing. ABGD analyses were performed on 94 datasets (one family, 12 subfamilies and 81 genera datasets) since datasets composed by less than three sequences were excluded from the analyses. The 3% nucleotide distance threshold analyses were carried out on 96 datasets (one family, 12 subfamilies and 83 genera datasets).

Molecular species delimitation analyses

Concerning ABGD, the average percentage of observed matches on all datasets was 77.6% (for analyses on family, subfamily and genera, respectively 69.2%, 79.2%, 77.5%); the average percentage of merge, mixture and split cases were respectively of 10.9%, 2.8% and 8.7% (Fig 2).

Delimitation analyses using the 3% nucleotide distance threshold resulted in an average percentage of 73.9% match (65.5%, 72.5%, 74.2% for family, subfamilies and genera datasets respectively), 5.1% merge, 4.4% mixture and 16.6% split cases.

Regarding GMYC analyses, the likelihood-ratio test rejected the null model for 43 over 94 datasets; for the datasets for which GMYC delimitation model was not accepted it was considered that no match, merge, split or mixture cases were observed. The average percentage of matches resulting from GMYC analyses on haplotype-collapsed DSs was 33.2% (62%, 61.9% and 28.6% for datasets at the level of family, subfamilies and genera, respectively). The average percentage of merge, mixture and split cases were respectively of the 3.6%, 4.3% and 12.2% (Fig 2).

Regarding mPTP analyses, for 90 over 94 datasets ASDDSV values resulted < 0.01, indicating that the ten independent MCMC runs converged. For the majority of the datasets (83 over 94) ASV values resulted high (median: 92.4%), suggesting that the ML solution is supported by the data. For 11 datasets for which ML delimitations had low support (ASV values lower then 50%), it was considered

that no match, merge, split or mixture cases were present. From mPTP analyses on all datasets the average percentages of 58.7% match (55.7%, 71,8%, 56.8% for family, subfamilies and genera datasets respectively), 4.6% merges, 8.4% mixture and 16.7% split cases were estimated (Fig 2).

Considering the agreement among methods, the four delimitation methods applied to family dataset delimit in the same category the 68.8% of the species: 50.4% as match, the 3% as merge, the 7.6% as mixture and the 7.9% as split. Regarding subfamilies datasets, the methods agreed in the delimitation of 71.1% of the species: 53.7% as match, 3.5% as merge, 7.4% as mixture and 6.5% as split, and for genera dataset the 52.7%: 42.6% as match, 1.8% as merge, 3.5% as mixture and 4.8% as split (Fig 2).

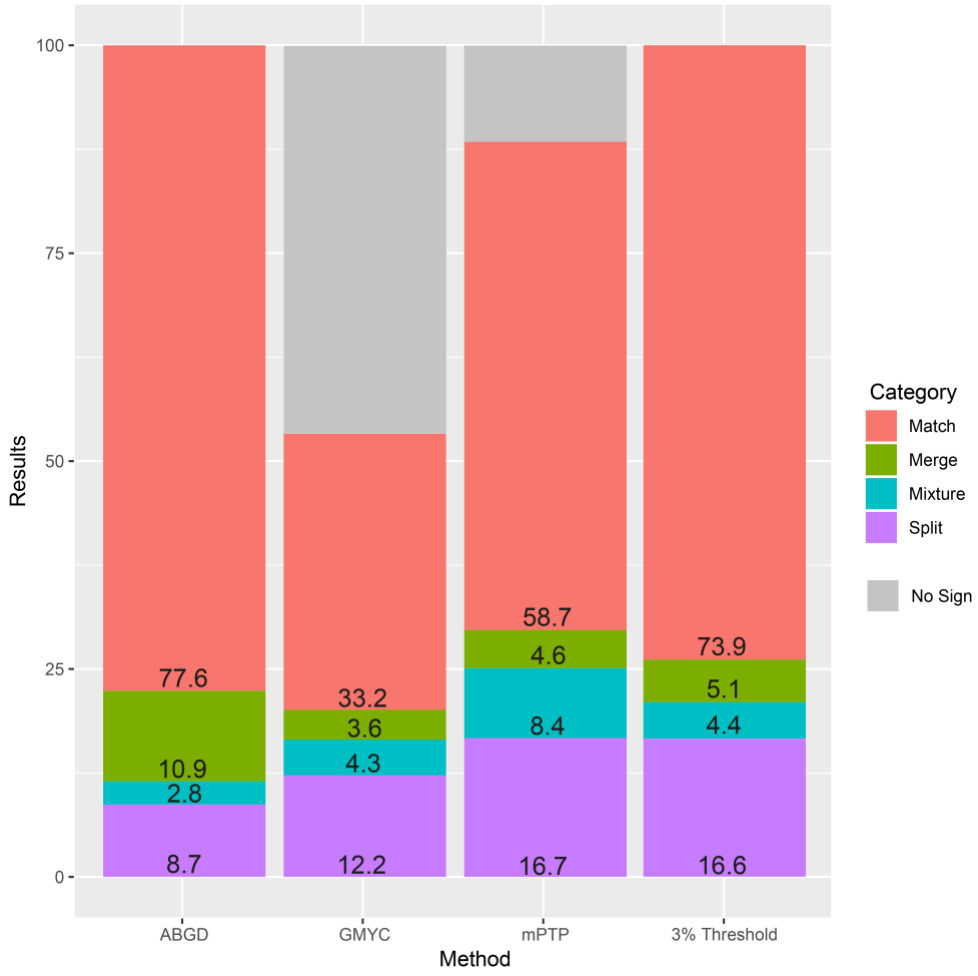

Figure 2. Molecular species delimitation results categorized basing on the accordance between molecular delimited units and morphological species.
The average of the results obtained on all datasets with each delimitation methods are expressed in percentage.

Species biology related factors

Mean number of haplotypes per dataset. Within the analysed datasets are present in mean ~6 haplotypes per species. In the family dataset the mean number of haplotypes per species is 6.71, in average in subfamilies and genera datasets are present respectively 5.55 (range: 2-13.5) and 5.35 (range: 1-40.5) haplotype per species; two genera datasets, *Macrocoma* and *Xanthogaleruca* (both composed by three sequences belonging to one morphological species) include all sequences of the same haplotype. A strong positive correlation was found between the number of sequences in the dataset and the *Mnhd* (r= 0.96).

Species morphological identification difficulty level. Among the species included in the family dataset, 35.1% were defined as easy to be morphologically identified (*dl* I), 53.4% were assigned to *dl* II and 11.5% to *dl* III. On average, within subfamilies datasets 65.2% of the species are categorize in the *dl* I, the 29% and 5.8% in the *dl* II and III respectively. Regarding genera datasets, the classification resulted as follow: in average 61.9% of the species assigned to *dl* I, 34.2% to *dl* II and 3.9% to *dl* III. Among the species categorized in *dl* I there are those of the subfamily datasets Zeugophorinae, Synetinae and Hispinae and those included in 38 out of the 83 genus datasets considered in this study; all the species included in 11 genera datasets were assigned to *dl* II (e.g. *Neocrepidodera*, *Orestia*, *Plagiosterna*, *Plateumaris*) and most of the species of *Altica* and *Oulema* datasets to *dl* III.

Study experimental design related factors

Median of the geographic distance between conspecific collection. In family dataset the *mgdc* is ~219 km; in average a median distance of 283 km [179-528 km] and 390 km [0-1685 km] were estimated for subfamilies and genera datasets. The higher *mgdc* were obtained for the genus dataset *Tituboea* including one species (*Tituboea biguttata*) and two specimens that were collected one in Italy and one in Morocco (1,685 km of distance), followed by *Lilioceris* (three species) including 24 specimens collected in Italy, Germany, Estonia and Finland (*mgdc* 1,528 km).

Number of morphological species included in each dataset. The family dataset is composed by 542 species, in subfamily and genus datasets there are in mean 45 morphospecies (range: 1-195) and 7 morphospecies (1-93) per dataset respectively. The most represented subfamily resulted to be that of flea beetles (Alticinae) with 195 species, followed by Cryptocephalinae (147 species), and the less represented Synetinae and Zeugophorinae (one species) in accordance with their abundance in the Euro-Mediterranean area. Regarding genera, *Cryptocephalus* (93 species) and *Longitarsus* (49 species) are more represented then others.

Factors affecting the molecular delimitation

The factors significantly influencing species delimitation efficiency resulted to be the *msdm* (molecular species delimitation method adopted), the *mgdc* (the median value of the geographic distance between conspecific collection localities) and *dl III* (the morphological identification difficulty level) (Table 1). Regarding *msdm*, distance based methods resulted related to a higher delimitation efficiency than coalescent tree based (Fig. 3). A high number of species assigned to *dl III* in the dataset resulted significantly related to a decrease of delimitation efficiency (Fig. 4). Finally, E and *mgdc* showed a weak positively relation (Fig. 5).

The linear model results slightly changed when fitted on different size datasets; interestingly, the 3% nucleotide distance threshold delimitation method resulted significantly related with a low delimitation efficiency when the size of datasets increase (Table 2); conversely, the geographic distance among sampling points was found not significantly related with delimitation efficiency when the models are fitted on datasets including more than 5 sequences (Table 2). The influence of the other variables on delimitation efficiency remained unchanged.

Table 1. Mixed linear model analysis results on all datasets.

Tested factors	***Chisq***	***Df***	***P-value***
Dl II	0.47	1	0.49
Dl III	6.15	1	0.013
Mgdc	4.8	1	0.02
Dtr	4	2	0.14
Msdm	127.03	3	<0.001
Mnhd	0.01	1	0.91
M	0.33	1	0.57

Dl II = second difficulty level of morphological species identification; *Dl III* = third difficulty level of morphological species identification; *Mgdc* = median value of the geographic distance between conspecific collection localities; *Dtr* = dataset taxonomic rank; *Msdm* = molecular species delimitation method adopted; *Mnhd* = mean number of haplotypes per dataset; *M* = number of morphospecies per dataset

Table 2. Summary of results of the mixed linear model analyses performed on different size datasets.

N seq	N datasets	Msdm	Dtr	M	Mnhd	Dl II	Dl III	Mgdc
>=5	324	***	-	-	-	.	**	-
>=10	274	***	-	-	-	-	**	-
>=20	210	***	-	-	-	-	**	-
>=40	154	***	-	-	-	-	*	-
>=80	108	*	-	-	-	-	.	-
>=160	56	.	-	-	-	-	*	-

N seq = number of sequences per dataset; N datasets = number of analysed datasets; *Msdm* = molecular species delimitation method adopted; *Dtr* = dataset taxonomic rank; *M* = number of morphospecies per dataset; *Mnhd* = mean number of haplotypes per dataset; *Dl II* = second difficulty level of morphological species identification; *Dl III* = third difficulty level of morphological species identification; *Mgdc* = median value of the geographic distance between conspecific collection localities; *** p-value <0.001; ** p-value < 0.01; * P-value <0.05; . P-value ~0.05; - not significant.

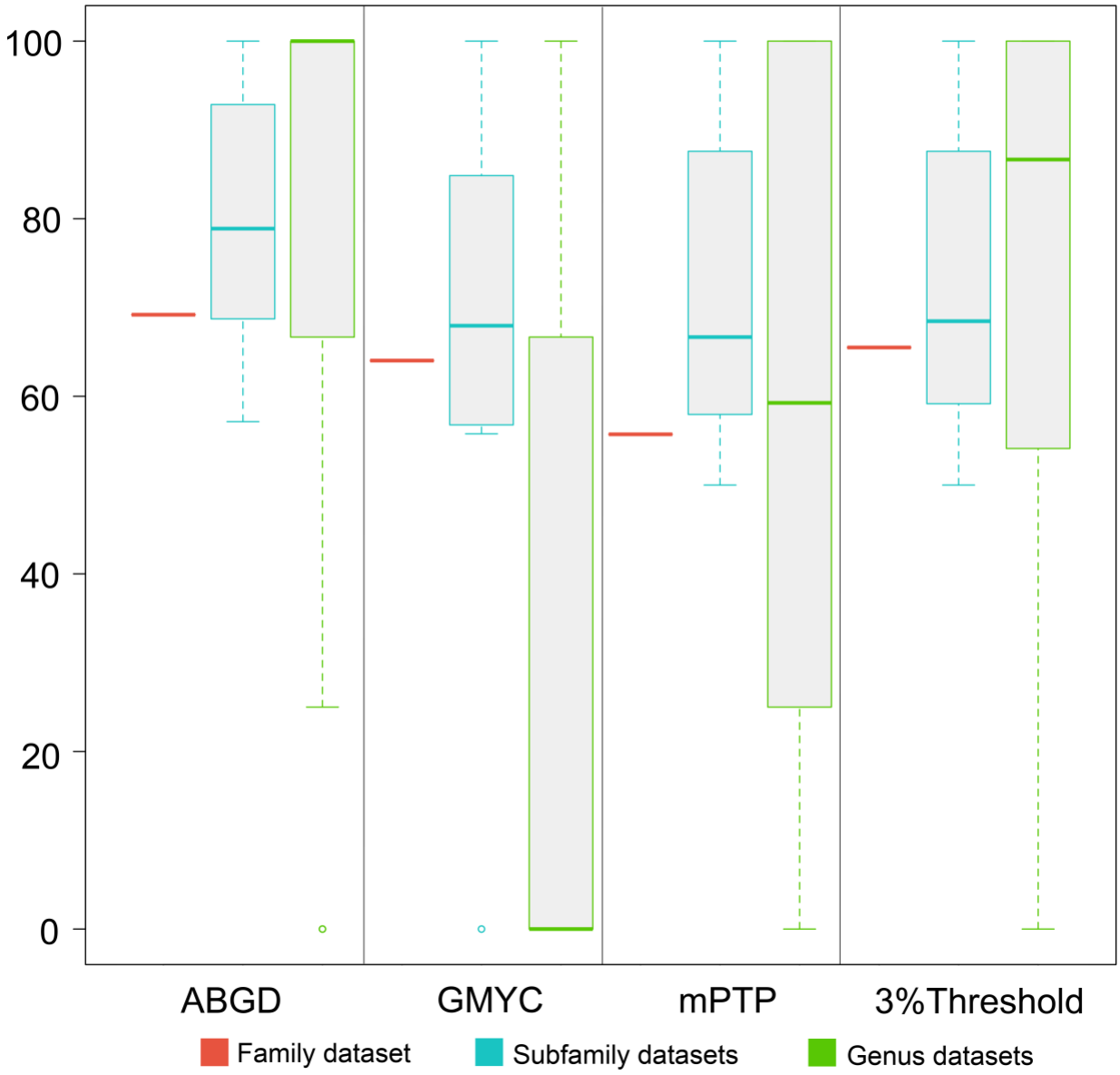

Figure 3. Distribution of matches for the three groups of analysed datasets calculated from the four molecular species delimitation methods results (ABGD, GMYC, mPTP and 3 % nucleotide distance threshold). The number of matches is expressed as percentage. In red the matches obtained from the family dataset analyses, in light blue those obtained from subfamily datasets analyses and in green from genus datasets analyses.

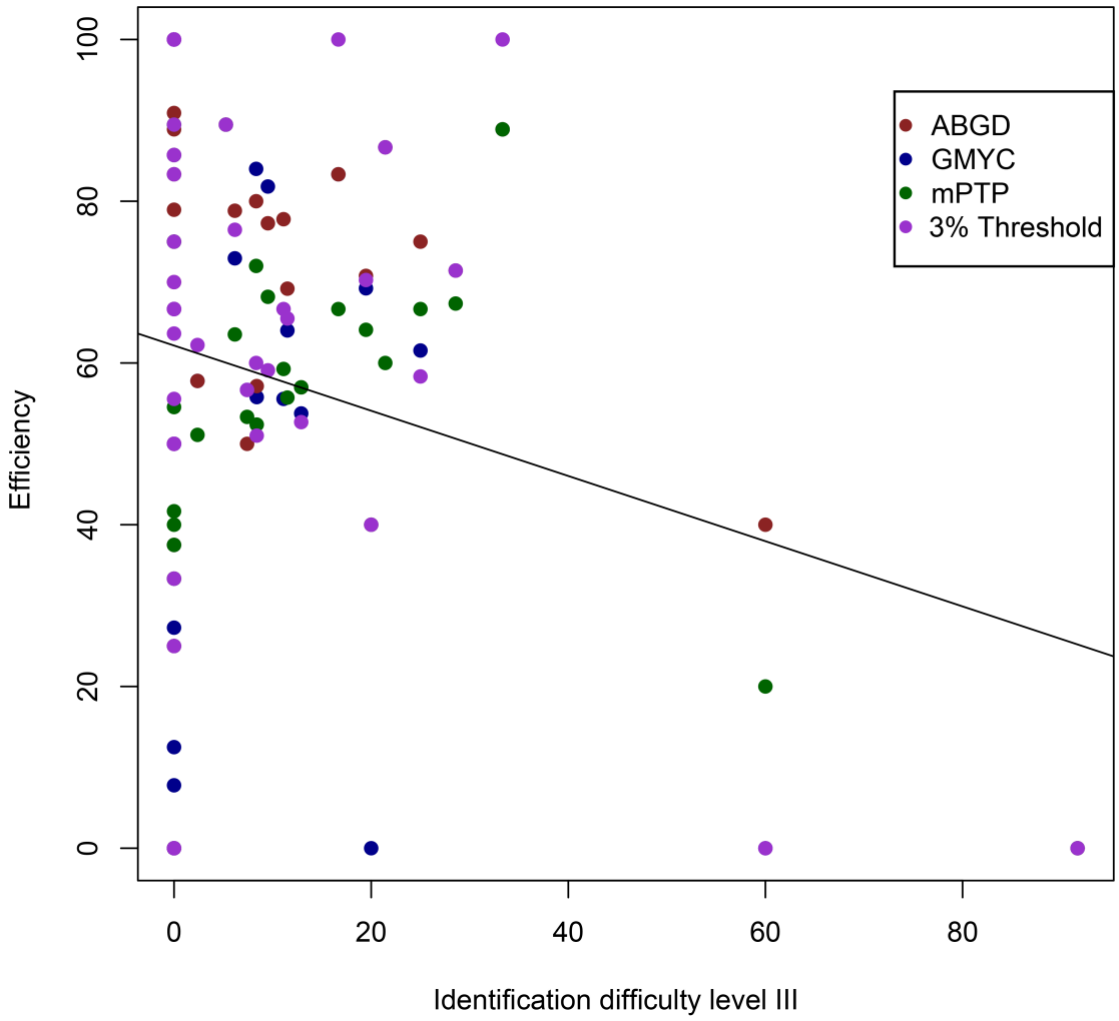

Figure 4. Effect of species morphological identification difficulty on molecular species delimitation methods efficiency. Brown dots represent ABGD results, navy blue dots GMYC results, green dots mPTP results and purple dots results of 3 % threshold analyses.

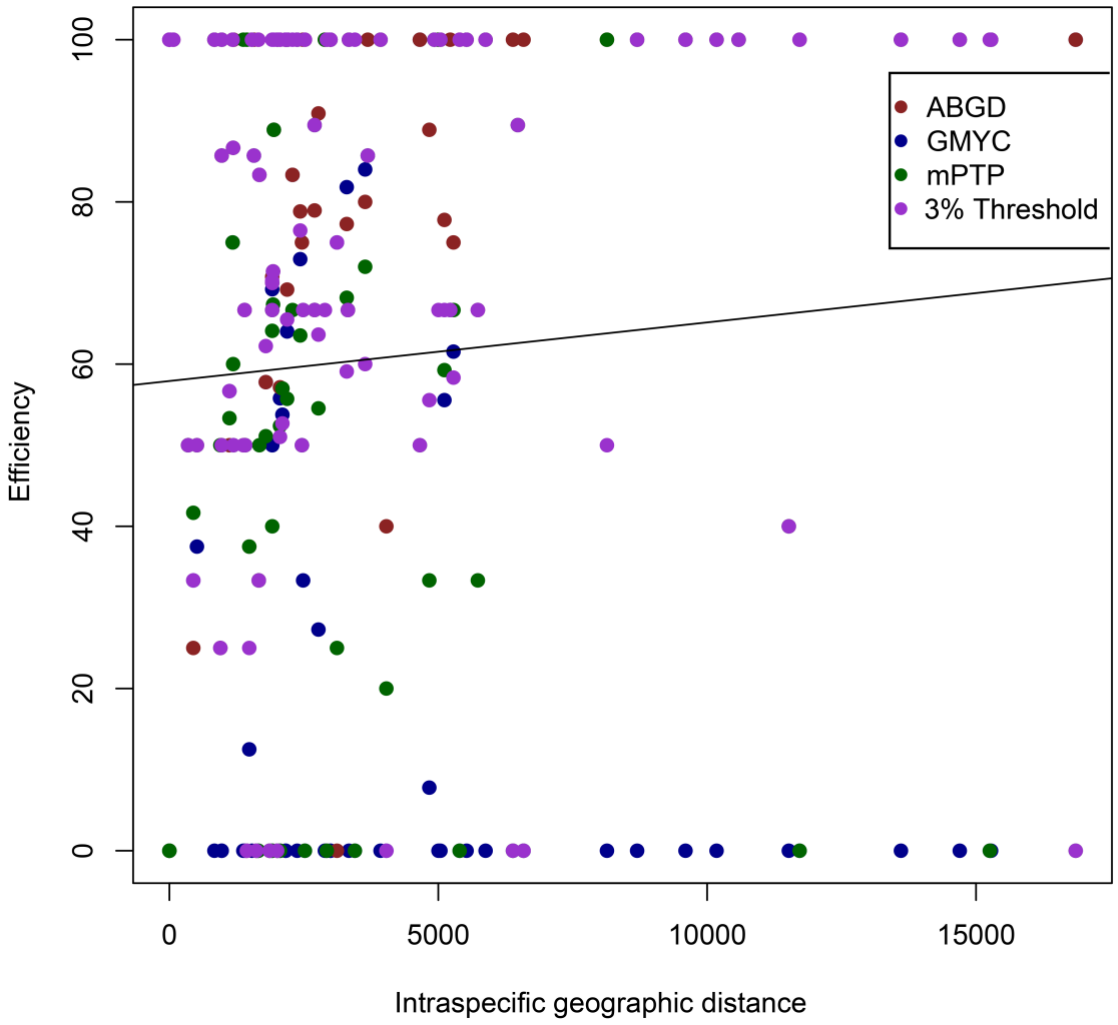

Figure 5. Effect of intraspecific geographic distance on molecular species delimitation methods efficiency. Brown dots represent ABGD results, navy blue dots GMYC results, green dots mPTP results and purple dots results of 3 % threshold analyses.

Discussion

In this work, the influence of some species biology and study experimental design related factors on molecular species delimitation efficiency was evaluated on more then 90 real datasets with different features. The obtained results provide suggestions useful to plan appropriate experimental design in species delimitation studies or be aware of the possible biases in the achieved results. Among the tested factors, those found to affect delimitation efficiency of all the methods used were i) the intraspecific geographic distance (*mgdc*), ii) the difficulty in species morphological identification (*msdm*) and iii) the

delimitation methods adopted (*dl III*). Hereafter, tested factors will be discussed.

Species biology related factors

Mean number of haplotypes per dataset. Something difficult to plan in a molecular delimitation study experimental design is how many haplotypes per species will be sampled. In any case, in this study a positive correlation between the number of sequences and the number of haplotypes per species was found (r=0.96) suggesting that the higher was the sampling effort, the higher was the number of sampled haplotypes.

In this work, the number of haplotypes per species was found to not significantly affect the delimitation efficiency. In previous studies an influence of this factor was observed; as example, in coalescent tree-based delimitation a low number of haplotypes per species was related to the oversplit of the entities, whose possible explanation is a not clearly observable transition between speciation and coalescent processes (Lohse, 2009); in this work no split was found when the number of haplotypes per species is low, neither for GMYC nor for mPTP. What it was observed is that, in the extreme cases in which the mean number of haplotypes per dataset (*mnhd*) was ≤ 2, the efficiency resulted higher for ABGD and 3% threshold delimitation methods (100% E on 12/12 datasets). This result is in accordance with the hypothesis that less haplotypes are analysed, lower is the probability to include intermediate haplotypes among species, and consequently more evident the barcoding gap (Meyer & Pauly, 2005). Interestingly, when the datasets were composed by 2 haplotypes of one species, the only case over 7 in which GMYC produced significant results was related to a high nucleotide divergence between haplotypes (*Tituboea* dataset, 6.6% mean intraspecific nucleotide distance, in comparison to the other datasets having intraspecific divergence <2%). This result is in accordance to Talavera et al., 2013, where was observed that GMYC species delimitation is not improved by a high number of haplotypes per species, but the delimitation obtained maintaining only the two most divergent haplotypes of a species has the same accuracy that having the intermediate haplotype among them.

Dataset taxonomic rank. In this work, no significant influence of the taxonomic rank of datasets on delimitation results was found. On family and subfamily datasets the used methods showed similar efficiency (55.7-79.2% E), the higher was observed for ABGD, while the lower for mPTP. On genus dataset coalescent tree-based methods showed the lower efficiencies (Fig. 3) and in particular for GMYC, many not significant analyses were obtained. In accordance to what observed by Talavera et al., 2013, this result can be related to the very low number of species included in some genera datasets, rather than to the taxonomic rank, but it is also dependent on the analysed. For example, on the genus

datasets *Altica* and *Oreina*, including 12 and 8 species respectively, the significance of GMYC delimitation was not reached.

Species morphological identification difficulty level. Three levels of morphological identification difficulty were defined for Chrysomelidae species and the highest difficulty level (*dl III*), was found to be related to a decrease of species delimitation efficiency. Only a small fraction of the Chrysomelidae species included in the analysed datasets was classified into the *dl III* (11.5%), which belong to closely related species groups, where different species are extremely similar in the morphology. Despite the morphological taxonomy of Chrysomelidae can be considered studied in detail and well defined, the identification of the species categorized in *dl III* is problematic also for specialists. Examples are some species included in the genus datasets *Galerucella*, *Chaetocnema* and *Longitarsus*; interestingly, these species where split or merged together by the delimitation methods. In these cases, species delimitation methods could have highlighted morphological identification errors that should not be regarded as the human operator inability, but as an evidence of one morphological taxonomy limit. This finding affirms the potential of the molecular species delimitation methods to assist morphological taxonomy in an integrative taxonomy framework. Another example of species classified in the *dl III*, is the species of the genus *Altica*; in the *Altica* genus dataset analysed in this study the 92% of the species were considered as belonging to *dl III* since different species are distinguishable only through subtitle differences in the shape of the aedeagus and spermatecha. In this case, all species delimitation methods, except GMYC for which the analysis was not significant, merged in one unit all the 12 analysed species; it suggests that neither morphology nor *cox1* based species delimitation, maybe due to incomplete lineage sorting or hybridisation phenomena, are able to clearly define species limits.

Study experimental design related factors

Impact of the number of species per dataset

The number of species per dataset (*m*) was found to not affect the delimitation efficiency, even though the datasets analysed in this work highly differ in the number of included species, ranging from 1 to 542. In other studies, nucleotide distance based methods efficiency was found to be affected by the increasing number of species; in particular, in large geographic scale studies, that usually include more species that regional scale ones, a lower delimitation efficiency using barcoding gap based methods was observed (Meyer & Pauly, 2005). This was found to be due to a higher probability of including closely related species that commonly show overlap in intra and interspecific nucleotide distance distribution, leading to the missing of the barcoding gap (Meyer & Pauly, 2005, Puillandre et al., 2012).

On our datasets, assembled taking into account taxonomic rank and not separated basing on the geographic area of specimens collection, closely related species are always analysed together so the previously reported effect was not observed. Regarding tree-based methods, some studies found an influence on GMYC results when the number of species is low; the number of significant GMYC analyses was observed to decrease when datasets are composed by less than 5 species (Talavera et al., 2013) and to be all not significant for datasets including 1 or 2 species maximum (Dellicour & Flot, 2015). The same pattern was seen in this work for GMYC and less frequently also for mPTP. GMYC analyses on 41 over 50 datasets including three or less than three species resulted not significant; in particular, all analyses on one ingroup species datasets resulted not significant, while four on two species dataset reach the significance. Despite performing species delimitation studies on three or less species is very rare, it could append when delimitation analyses aims to disentangle the relation among taxa of genera really poor in term of species; in this case, if it is not possible to increase the species sampling including the most closely related genus species, it is recommended to use distance based species delimitation methods.

Impact of intraspecific geographic distance

In this work, a positive but weak relation between methods efficiency and intraspecific geographic distance was found (Fig. 5). The weakness of this relation is confirmed also by the fact that when the mixed linear model analysis is repeated on > 5 sequences datasets (t_1) the significance is lost (Table 2). From species delimitation results it is possible to observe that a higher methods efficiency is found for those datasets for which the high intraspecific geographic distance is accompanied by low intraspecific nucleotide divergence. Some examples are *Arrhenocoela* (*mgdc* 1360 km, median intraspecific nucleotide distance 0.4%) *Agelastica* (*mgdc* 1470 km, median intraspecific nucleotide distance 0.3%), *Pyrrhalta* (*mgdc* 1526 km, median intraspecific nucleotide distance 0.3%), *Lilioceris* (*mgdc* 1528 km, median intraspecific nucleotide distance 1.9%). Distance based methods resulted in 100% E for all these datasets, mPTP had 100% E on *Arrhenocoela*, *Lilioceris* and *Agelastica* datasets and GMYC did not reach the delimitation significance for all datasets. A low methods efficiency in presence of low intraspecific geographic distance is observed for example for the datasets *Hydrothassa* (*mgdc* 347 km), *Pachybrachis* (*mgdc* 446 km), *Galeruca* (*mgdc* 512 km) where a high number of merges is observed.

Since isolation by distance (IBD) was found to be a very common phenomenon within phytophagous insects (Peterson & Denno, 1998), in this work we expected to observe that high intraspecific

geographic distances were related to high intraspecific nucleotide distances and consequently a higher tendency of the methods to split species. The different pattern found, make us to tentatively hypothesise that IBD phenomenon is not so spread among the analysed species and that in absence of it, the methods have high efficiency also when intraspecific nucleotide distances are high. Since the datasets analysed in this work are very variable in term of intraspecific sampling (range 1 to 238 specimens per species) and species distribution range coverage, and some of them are too incomplete for testing for the presence of some phenomena that shape intraspecific variability (IBD; isolation by environment, IBE), more data are required for performing appropriate tests and understand the influence of these phenomena on molecular species delimitation.

Impact of used method

In this work, the method used resulted to influence the delimitation efficiency. Among the species delimitation methods tested, the higher efficiency (E) values resulted for distance-based methods analyses (Fig. 3). ABGD average E was 77.6%, followed by 3% threshold (73.9% E), while coalescent tree-based methods showed lower E values (GMYC 33.2% E, mPTP 58.7% E); the lower efficiency of the latest methods is partially related to the lack of significance of some datasets analyses (43 for GMYC and 11 for mPTP; Fig. 2). Our result is consistent with another study comparing species delimitation methods performances (BIN, ABGD, GMYC and PTP) on a dataset of 5290 *cox1* sequences belonging to 1870 species of beetles, where the higher number of matches was observed for ABGD (Pentinsaari et al., 2017); moreover, in accordance with what previously observed (Lin et al. 2015; Pentinsaari et al., 2017), on our data this method gave the most conservative results, with the higher number of merges (10.9%) and the lower of splits (8.7%). Despite GMYC is known to be prone to oversplit species (Paz A & Crawford, 2012; Hamilton et al., 2014), in our work it was not associated with a high number of splits (12.2%) in comparison to the ~17% for mPTP and the 3% threshold (Fig. 2). Despite the 3% delimitation threshold used is higher than the value estimated to be the optimal threshold for Chrysomelidae distance based identification (1%, Magoga et al., 2018), a consistent number of splits was observed; this is due to the fact that within Chrysomelidae there are some taxa with a higher intraspecific variability then others, i.e. Cassidinae (optimal threshold 5.9%; Magoga et al., 2018) for which in this work the 40% of the species were split in the delimitation using 3% threshold. This make impossible to find a barcoding gap in the distribution of intra-interspecific nucleotide distances of the whole family, but also for lower taxonomic levels. For example, for Cryptocephalinae subfamily, the optimal threshold value was found to be 1%; this threshold is low

because within the subfamily, more than in others, are present a lot of closely related species groups characterized by low interspecific distances (Montagna et al., 2017; Magoga et al., 2018), but also other species with an intraspecific variability higher than the 5% (Magoga et al., 2018). Also for this subfamily our delimitation analysis using 3% threshold led to a consistent number of splits (22%). The fact that for Chrysomelidae also at lower taxonomic levels a barcoding gap cannot be found, make clear how a perfect threshold for any taxonomic level could not be estimate. Despite of this, the species delimitation using distance threshold had the second highest efficiency after ABGD.

Conclusions

In this work, a higher efficiency of molecular species delimitation distance based methods was observed, especially when the sampling size is low; frequently when datasets including a very low number of species and/or haplotypes per species were analysed through coalescent tree-based methods, in particular GMYC, the delimitation model did not result to be better than the null model, and consequently the significance of the analysis was not reached. For this reason, we are prone to suggest the use of distance-based methods in these cases. Moreover, the finding that the methods efficiency decreases when species very difficult to be morphologically identified are delimited, highlights the limit of morphological taxonomy and makes prone to suggest the use of integrative taxonomy to overcome this limit. Despite distance-based methods are not taking in account the evolutionary history of species and the barcoding gap is not an unerring principle for finding the limit between intraspecific and interspecific level, among the tested methods it has been found to give the delimitation more consistent with the one of morphological approach.

2.5.3 References

Biondi, M., Urbani, F. & D'Alessandro, P. (2013) Endemism patterns in the Italian leaf beetle fauna (Coleoptera, Chrysomelidae). *ZooKeys* **332:** 177-205.

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? In: Schmitt, Chaboo, Biondi (Eds) Research on Chrysomelidae 8. *ZooKeys* **856:** 3-25.

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.

Canty, R., Ruzzier, E., Cronk, Q., & Percy, D. (2016). *Salix* transect of Europe: patterns in the most abundant chrysomelid beetle (Coleoptera: Chrysomelidae) herbivores of willow from Greece to Arctic Norway. *Biodiversity data journal* **4:** e10194.

Carstens, B. C., Pelletier, T. A., Reid, N. M. & Satler, J. D. (2013) How to fail at species delimitation. *Molecular Ecology* **22:** 4369-4383.

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:** 485-495.

Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature methods* **9:** 772.

Davis, J. I. & Nixon, K. C. (1992) Populations, genetic variation, and the delimitation of phylogenetic species. *Systematic Biology* **41:** 421-435.

Dayrat, B. (2005) Towards integrative taxonomy. *Biological Journal of the Linnean Society* **85:** 407-417.

Dellicour, S. & Flot, J.-F. (2015) Delimiting Species-Poor Data Sets using Single Molecular Markers: A Study of Barcode Gaps, Haplowebs and GMYC. *Systematic biology* **64:** 900-8.

Dellicour, S. & Flot, J. (2018) The hitchhiker's guide to single-locus species delimitation. *Molecular ecology resources* **18:** 1234-1246.

deWaard, J. R., Levesque-Beaudin, V., deWaard, S. L., Ivanova, N. V., McKeown, J. T. A., Miskie, R., Naik, S., Perez, K. H. J., Ratnasingham, S., Sobel, C. N., Sones, J. E., Steinke, C., Telfer, A. C., Young, A. D., Young, M. R., Zakharov, E. V. & Hebert, P. D. N. (2019) Expedited assessment of terrestrial arthropod diversity by coupling Malaise traps with DNA barcoding. *Genome* **62:** 85-95.

Dincă, V., Montagud, S., Talavera, G., Hernández-Roldán, J., Munguira, M. L., García-Barros, E., Hebert, P. D. N & Vila, R. (2015) DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. *Scientific Reports* **5**: 12395.

Doguet, S. (1994) Coléoptères chrysomelidae: Alticinae. Volume 2. Federation francaise des societes de sciences naturelles. 694 pp.

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:** 1969-1973.

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:** 4422-4436.

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:** 3678-3686.

Ezard, T., Fujisawa, T. & Barraclough, T. G. (2009) splits: SPecies' LImits by Threshold Statistics. R package version.

Fontaneto, D., Herniou, E. A., Boschetti, C., Caprioli, M., Melone, G., Ricci, C. & Barraclough, T. G. (2007) Independently evolving species in asexual bdelloid rotifers. *PLoS Biology* **5:** e87.

Fox, J. & Weisberg, S. (2019) An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

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:** 707-724.

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:** e0216786.

Gómez-Zurita, J., Cardoso, A., Coronado, I., De la Cadena, G., Jurado-Rivera, J. A., Maes, J. M., Montelongo, T., Nguyen, D. & 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.

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:** 50-56.

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:** 79-93.
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:** 313-321.
Hebert, P. D. N. & Gregory, T. R. (2005) The Promise of DNA Barcoding for Taxonomy. *Systematic Biology* **54:** 852-859.
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:** 729-739.
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 The Biology of Plethodontid Luo, A., Ling, C., Ho, S. Y. W., Zhu, C.-D. Comparison of Methods for Molecular Species Delimitation Across a Range of Speciation Scenarios. *Systematic Biology* **67**: 830-846.
Hijmans, R. J. (2017) Geosphere: Spherical Trigonometry. R package version 1.5-7. https://CRAN.R-project.org/package=geosphere.
Hudson, R. R. (1990) Gene genealogies and coalescent process. Futuyma D.J., Antonivics J. (eds). Oxford Surveys in Evolutionary Biology. Oxford: Oxford University Press. p. 1-44.
Jolivet, P. (2015) Together with 30 years of Symposia on Chrysomelidae! Memories and personal reflections on what we know more about leaf beetles. *ZooKeys* **547:** 35-61.
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:** 1630-1638.
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.
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:** 1547-1549.
Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2017) lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* **82:** 1-26.
Lin, X., Stur, E. & Ekrem, T. (2015) Exploring Genetic Divergence in a Species-Rich Insect Genus Using 2790 DNA Barcodes. *PLoS ONE* **10:** e0138993.
Lohse, K. (2009) Can mtDNA barcodes be used to delimit species? A response to Pons et al. (2006). *Systematic Biology* **8:** 439-442.
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.
Meyer, C. P. & Paulay, G. (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology* **3:** 2229-2238.
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:** e0199609.
Monaghan, M. T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D. J. G., Lees, D. C., Ranaivosolo, R., Eggleton, P., Barraclough, T. G. & Vogler, A. P. (2009) Accelerated Species Inventory on Madagascar Using Coalescent-Based Models of Species Delineation. *Systematic Biology* **58:** 298-311.
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:** 243-253.
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.
Müller, G. (1953) I Coleotteri della Venezia Giulia. II. Coleoptera Phytophaga. Trieste, 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.
Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20:** 289-290.

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:** 887-896.

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:** 393-404.

Peterson, M. A. & Denno, R. F. (1998) The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. *The American Naturalist* **152:** 428-46.

Plewa, R., Sikora, K., Gutowski, J. M., Jaworski, T., Tarwacki, G., Tkaczyk, M., Rossa, R., Hilszczanski, J., Magoga, G. & Kajtoch, Ł. (2018) Morphology, genetics and *Wolbachia* endosymbionts support distinctiveness of *Monochamus sartor sartor* and *M. s. urussovii* (Coleoptera: Cerambycidae). *Arthropod Systematics & Phylogeny* **76:** 123-135.

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D. & Vogler, A. P. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* **55:** 595-609.

Porter, A. H. (1990) Testing nominal species boundaries using gene flow statistics: the taxonomy of two hybridizing admiral butterflies (Limenitis: Nymphalidae). *Systematic Zoology* **39:** 131-147.

Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G. (2012) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* **21:** 1864-1877.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reid, N. M. & Carstens, B. C. (2012) Phylogenetic estimation error can decrease the accuracy species delimitation: a Bayesian implementation of the General Mixed Yule Coalescent model. *BMC Evolutionary Biology* **12:** 196.

Revelle, W. (2018) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 1.8.12.

Rubinoff, D. & Holland, B.S. (2005) Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic Biology* **54:** 952-961.

Ruppert, K. M., Kline, R. J. & Rahman, M. S. (2019) Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global Edna. *Global Ecology and Conservation* **17:** e00547.

Swartz, E. R., Mwale, M. & Hanner, R. (2008) A role for barcoding in the study of African fish diversity and conservation. *South African Journal of Science* **104:** 293-298.

Talavera, G., Dinca, V. & Vila, R. (2013) Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. *Methods in Ecology and Evolution* **4:** 1101-1110.

Telfer, A. C., Young, M. R., Quinn, J., Perez, K., Sobel, C. N., Sones, J. E. et al. (2015) Biodiversity inventories in high gear: DNA barcoding facilitates a rapid biotic survey of a temperate nature reserve. *Biodiversity data journal* **3:** e6313.

Thomsen, P. F. & Willerslev, E. (2015) Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity, *Biological Conservation* **183:** 4-18.

Warchalowski, A. (2003) Chrysomelidae. The leaf-beetles of Europe and the Mediterranean area. Warszawa: Nautura Optima Dux Fundation.

Weitschek, E., Van Velzen, R., Felici, G. & Bertolazzi, P. (2013) BLOG 2.0: a software system for character-based species classification with DNA Barcode sequences. What it does, how to use it. *Molecular Ecology Resources* **13:** 1043-6.

Wright, S. (1978) Evolution and the Genetics of Population, Variability Within and Among Natural Populations. The University of Chicago Press, Chicago.

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:** 402-420.

Zhang, J., Kapli, P., Pavlidis, P. & Stamatakis, A. (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* **29:** 2869-2876.

2.5.4 Personal contribution to the work

Coinceving of the study in collaboration with M.M. and D.F. Bioinformatics and statistical analyses, writing of the manuscript draft.

3. Conclusions

In this thesis, the study of insects taxonomy through molecular approaches has been explored and its efficiency evaluated. DNA-barcoding is confirmed as a suitable tool for leaf beetles (Coleoptera, Chrysomelidae) identification, since high efficient in the identification of the species belonging to this family (chapters 2.1 and 2.2 of this thesis). Beyond the benefit of using DNA-barcoding for the study of the taxonomy of Chrysomelidae (chapter 2.1), the applicability of this tool for improving the monitoring of pest species has been demonstrated (chapter 2.2).

DNA-based species delimitation methods, employed in an integrative taxonomic framework, have proved to be an information source that frequently allows overcoming the limits related to the use of morphology alone in species delimitation (chapters 2.3 and 2.4). Despite the usefulness of molecular species delimitation methods, it has been shown how their efficiency can decrease in some conditions (chapter 2.5), suggesting that more efforts should be done for clarifying how to make the best possible use of them. The study of Chrysomelidae through molecular taxonomy performed during my PhD has led to the discovery of some cases of incogruence between morphological and molecular signal that led to undertake some studies not included in this thesis. Among them an integrative taxonomy species delimitation study on west Palearctic *Phyllotreta* species, in order to determine the presence of undescribed diversity, with a particular focus of *Phyllotreta nigripes*, a Brassicaceae pest species that seems to include a cryptic species. Another integrative taxonomy species delimitation study has been started in order to desentangle the relationships among the taxa included in the *Cryptocephalus marginellus* species group. Moreover, thanks to the evidences found during Euro-Mediterranean Chrysomelidae barcode dataset development a taxonomic revision of *Altica* genus, also here integrating molecular, morphological and ecological information about the species of the genus, is going to start soon. Finally, the knowlege achieved during my thesis has proven useful for the development of a dataset of barcode sequences for exploring the poorly studied leaf beetle fauna of Iran (ongoing project).