Habitat selection, cryptic diversity, phylogeny, and phylogeography of the European *Lepidocyrtus lanuginosus* species group (Collembola: Entomobryidae)

Dissertation

for the award of the degree

Ph.D. Division of Mathematics and Natural Sciences

of the Georg-August-Universität Göttingen

within the doctoral program Biology

of the Georg-August University School of Science (GAUSS)

submitted by

Bing Zhang

from

Zaoyang, Xiangyang, Hubei province, China

Göttingen, November 2018

Thesis Committee

Stefan Scheu, Animal Ecology, Johann-Friedrich-Blumenbach Institute for Zoology and Anthropology

Mark Maraun, Animal Ecology, Johann-Friedrich-Blumenbach Institute for Zoology and Anthropology

Members of the Examination Board

Reviewer: Stefan Scheu, Animal Ecology, Johann-Friedrich-Blumenbach Institute for Zoology and Anthropology

Second Reviewer: Christoph Bleidorn, Animal Evolution and Biodiversity, Johann-Friedrich-Blumenbach Institute for Zoology and Anthropology

Additional Reviewer (if applicable):

Further members of the Examination Board:

Elvira Hörandl, Systematics, Biodiversity and Evolution of Plants (with Herbarium), Albrecht-von-Haller Institute for Plant Sciences

Klaus Hövemeyer, Animal Ecology, Johann-Friedrich-Blumenbach Institute for Zoology and Anthropology

Mark Maraun, Animal Ecology, Johann-Friedrich-Blumenbach Institute for Zoology and Anthropology

Marko Rohlfs, Population and Evolutionary Ecology Group, Institute of Ecology, University of Bremen

Date of the oral examination: 14th of December, 2018

Table of contents

Summary

Accurate species identification and assessment of species diversity are essential for studies on phylogeny and phylogeography, adaptation and ecological function. The development of molecular methods triggered the discovery of cryptic species, i.e., genetically distinct lineages in morphologically undifferentiated species. Collembola (Arthropoda, Hexapoda) are one of the most numerous soil-living animals occurring in virtually all terrestrial ecosystems and habitats. Species delimitation is particularly difficult in Collembola due to considerable morphological conservatism, and many Collembola species comprise high genetic divergence and high cryptic species diversity. DNA-based methods provide useful tools for species delimitation, phylogenetic reconstruction, and lineage divergence time estimation. By analyzing two mitochondrial and two nuclear genes from three morphospecies of the European *Lepidocyrtus lanuginosus* species group (Collembola: Entomobryidae) from different geographic regions of Europe, this thesis focuses on exploring cryptic species / lineages diversity, their phylogeny, and the effects of historical geographic and climatic changes on the divergence and distribution of this species group in Europe.

In **chapter II**, I investigated phylogenetic relationships and genetic distances between populations of the morphospecies *L. lanuginosus* (Gmelin, 1788) from three different habitats in Central Europe, i.e. arable fields, grasslands, and forests replicated at six locations. Geographic distances between sampling locations were considerably larger than between habitat types. All four genes clearly separated the morphospecies *L. lanuginosus* into three major genetic lineages, with one of these lineages being close to *Lepidocyrtus cyaneus* Tullberg, 1871. The three lineages were genetically as distant to each other as well separated species. Selective colonization of the three habitats by these lineages indicate that they are sorted by habitats: one lineage was common and occurred in each of the three habitat types but preferentially in arable land; the second was restricted to forest; the third, although rare, preferentially occurred in grassland. The results indicate that genetic markers are a reliable and fast method to detect cryptic species, which may facilitate taxonomic research on Collembola species and the identification of possible species-specific morphological characters.

In **Chapter III**, I delimited species boundaries of the *L. lanuginosus* species group sampled across Central and Southern Europe (north and south of the Alps) by utilizing three DNA-based methods,

ABGD, PTP, and BPP. Species diversity delimited by morphology was compared with that delimited by genes. Three methods based on mitochondrial COI and COII congruently identified ten and nine distinct genetic lineages in the morphospecies *L. cyaneus* and *L. lanuginosus*, respectively. ABGD delimitated species barcoding gaps with K2P distances of 0.055–0.095 and 0.06–0.115 for COI and COII, respectively, within the species group. EF1-α separated 89% of these lineages, showing a higher resolution than 28S rDNA D1–2 in distinguishing closely related genetic lineages of the species group. The phylogenetic analysis based on the four genes showed that both morphospecies *L. cyaneus* and *L. lanuginosus* are polyphyletic, suggesting that body color is insufficient for delimiting morphospecies and lineages in this Collembola species group. This study challenges the current morphology-based species delimitation in the *L. lanuginosus* species group and suggests that molecular approaches are needed for accurate determination of Collembola species in both taxonomic and ecological studies. Overall, the results suggest that wide geographic sampling combined with molecular phylogenetic approaches is necessary to delimit species, understand the full range of cryptic diversity, and analyze phylogenetic relationships in Collembola.

In **Chapter IV**, I studied the phylogeography of 18 lineages of the *L. lanuginosus* species group using a multi-locus molecular approach. The genetic diversity and population structure of all lineages were analyzed using COII, while all four above-mentioned genes were concatenated to infer the phylogeographic origin of these lineages. Results showed that the 18 lineages did not overlap in their distribution ranges in Central, Southern, and Southeastern Europe, suggesting high genetic structure and limited gene flow between these three regions. The major lineages diverged in the Late Miocene and Pliocene (17–2.59 million years ago, Mya), i.e. before Quaternary ice ages, indicating that distinct lineages survived in multiple refugia in each sampling region during Quaternary glacial periods. The genetic structure of the 18 lineages of the group supported a model of sequential allopatric diversification within each sampling region. Further, three distinct lineages which diverged during the Pleistocene and Holocene were widely distributed across Central Europe, suggesting that glacial cycles in the Quaternary affected the spread of these lineages. Identical haplotypes of both of these lineages, occurring in localities hundreds of kilometers apart, suggest recent human-mediated dispersal across Central Europe. These results indicate that distribution patterns of Collembola in Europe are more complex than previously assumed.

By utilizing DNA-based analyses, my thesis highlights a novel view of the diversity, ecology, phylogeny, and phylogeography of the three morphospecies of the European *L. lanuginosus* species group. DNA-based analyses rejected the three-species hypothesis based on the morphological species concept and the monophyly of each species based on the phylogenetic species concept. This suggests that historical geographic and climatic changes dating to the late Miocene as well as recent human-mediated dispersal caused the lineage divergence and shaped the present-day distribution of this species group. Environmental factors other than geographic distances likely impede gene flow among lineages. Overall, the results of this thesis suggest that soil animals likely experienced different and more complex evolutionary forces than aboveground animals and plants in shaping their genetic diversification and biogeographic distribution. Future studies need to explore the physiological characters responsible for habitat sorting of cryptic species / lineages of the *L. lanuginosus* species group, and global scale phylogeographic studies on a number of Collembola species are needed for a deeper understanding of the dispersal, speciation, and evolution of Collembola and of soil animals in general.

Chapter I

General Introduction

The basis for a wide range of biological studies is the accurate identification and assessment of species diversity. Collembola (Arthropoda, Hexapoda), commonly referred to as springtails, are one of the most numerous soil-living animals, occurring in virtually all terrestrial ecosystems and habitats (Hopkin, 1997). The taxonomical categories of Collembola are entirely based on external morphology, with about 8,800 Collembola species currently described worldwide (Bellinger et al., 1996-2018; Deharveng, 2004). Due to considerable morphological conservatism, species delimitation is particularly difficult in Collembola. Morphological approaches are limited in disentangling species clusters, delimiting sibling species and color pattern forms, and reconstructing phylogeny. Molecular tools are extremely useful in evaluating the tenuous morphological characters, reconstructing highest relationships, and testing the accuracy of current taxonomical hierarchy in Collembola (D'Haese, 2002; Luan et al., 2005; Soto-Adames, 2000; Xiong et al., 2008; Zhang et al., 2015, 2014a; Zhang and Deharveng, 2015).

This thesis focuses on revealing habitat selection, cryptic diversity, phylogeny, and phylogeography of three morphospecies of the *Lepidocyrtus lanuginosus* species group (Collembola: Entomobryidae) by analyzing genetic diversity and population structure across Europe. High intra-specific genetic divergence and non-monophyly of the *L. lanuginosus* species group (Cicconardi et al., 2010; Mateos et al., 2018) raise concerns about the traditional morphological species delimitation, phylogenetic reconstruction, species distribution, and ecology of the species group in Europe. Molecular studies reveal that some morphology-based species comprise high numbers of cryptic species, i.e., genetically distinct lineages exist in morphologically undifferentiated species (Cicconardi et al., 2013, 2010; Porco et al., 2012a,b; von Saltzwedel et al., 2017). Two morphospecies of the *L. lanuginosus* species group, i.e., *L. cyaneus* Tullberg, 1871 and *L. lanuginosus* (Gmelin, 1788), are widely distributed and intensively studied in Europe and comprise high intra-specific genetic divergence and few morphological variations (Cicconardi et al., 2010; Mateos, 2012). Detailed genetic analyses of populations from a large sampling region are expected to allow inferring accurate species identification and assessment of the species group, providing genetic insight into diversity, distribution and speciation of Collembola.

DNA-based approaches have been developed to improve the evaluation of cryptic species diversity (Knowles and Carstens, 2007; Wiens, 2007), and a combination of relatively variable mitochondrial genes and conserved nuclear genes have been proven to be efficient for reconstructing

a reliable phylogeny and analyzing the phylogeography of Collembola species (von Saltzwedel et al., 2016; Yu et al., 2017). In this thesis, I first investigated the genetic divergence of the morphospecies *L. lanuginosus* at a local scale in Central Europe, to test whether cryptic species / lineages exist and whether these lineages have similar distribution pattern in different habitats (**Chapter II**). After confirming the existence of distinct genetic lineages in *L. lanuginosus*, I investigated the genetic diversity and phylogenetic relationships of cryptic species / lineages of the *L. lanuginosus* species group across Europe (**Chapter III**). I further analyzed the geographic distribution of the lineages in Central, Southern, and Southwestern Europe, and estimated their divergence times to investigate how past climate and geographic change affected their present-day distributions in Europe (**Chapter IV**). The overall goal of this thesis is to reveal the genetic structure, diversity, distribution and diversification patterns of two well-described, widely distributed, and intensively studied Collembola morphospecies. This will help identify the factors driving dispersal, diversification, evolution, and speciation in Collembola.

1.1 Genetic markers for molecular delimitation, phylogeny, and phylogeography

To achieve these aims, genetic sequences were integrated with morphological and biogeographic data. I sequenced two variable mitochondrial genes, i.e., cytochrome c oxidase subunit I and II (COI and COII), and two conserved nuclear genes, i.e., 28S ribosomal DNA D1–2 domain (28S D1–2), and Elongation Factor 1-α (EF1-α). COI (Hebert et al., 2003; Hogg and Hebert, 2004) is very efficient in molecular delimitation of Collembola (Katz et al., 2015; Sun et al., 2017; Yu et al., 2017; F. Zhang et al., 2018). COII has been used to study the population genetic divergence and phylogeny of Collembola, including *Desoria* (Stevens et al., 2007), *Isotoma* (Frati and Carapelli, 1999), *Lepidocyrtus* (Cicconardi et al., 2010; Mateos et al., 2018), Neanuridae (Frati et al., 2000b; Frati and Dell'Ampio, 2000), and *Orchesella* (Frati et al., 2000a, 1997; Timmermans et al., 2005). Nuclear 28S rDNA and EF1-α are efficient in discriminating Collembola species (Anslan and Tedersoo, 2015; Cicconardi et al., 2010; Zhang et al., 2014b) and for reconstructing the phylogeny of Collembola at genus and species level (Cicconardi et al., 2013, 2010; Mateos et al., 2018; Zhang et al., 2014b). Thus in **Chapter II**, all four genes were used as barcoding markers to calculate genetic distances within and between different species of *Lepidocyrtus*, and in **Chapter III**, the COI data set was used to calculate the barcoding gaps and to delimit species boundaries within the *L. lanuginosus* species group. Further, the efficiency of COII, 28S rDNA and EF1-α as alternative barcoding markers was evaluated by analyzing the congruency of cryptic species identification based on each of three genes compared to COI.

Various molecular markers, e.g., 18S and 28S ribosomal DNA (rDNA) and EF1-α, mitochondrial COI and COII and 16S rRNA, and even mitochondrial genome data (Carapelli et al., 2014), have been used to explore the phylogenetic relationships among Collembola. Among them, 18S and 28S rDNA are relatively conserved and were used for studying phylogeny at high taxonomic levels, i.e. among orders and families (D'Haese, 2002; Dell'Ampio et al., 2002; Luan et al., 2005; Xiong et al., 2008; Yu et al., 2016; Zhang et al., 2014a). Mitochondrial genes, e.g., COI, COII, and 16S, are more variable as compared to the above mentioned nuclear genes, and thus were combined with 18S, 28S rDNA or EF1-α to study phylogenies within genera of Collembola, for example in *Coecobrya* (F. Zhang et al., 2018), *Cryptopygus* (McGaughran et al., 2010a), *Entomobrya* (Ding et al., 2018), *Lepidocyrtus* (Cicconardi et al., 2013, 2010; Mateos et al., 2018), and *Megalothorax* (Schneider and D'Haese, 2013). The phylogenetic trees reconstructed with the concatenated data set of COI, COII, 28S D1–2 and EF1-α gave the highest support values, thus the concatenated data set was used for phylogenetic reconstructions in all three chapters.

Single mitochondrial genes such as COI and COII have been used to study the genetic divergence within Collembola species (Beet et al., 2016; Stevens et al., 2007). However, as mentioned above, some widely distributed Collembola species comprise cryptic species / lineages with similarly high intra- and inter-specific genetic divergences. Genetic divergence between species within genera reached 10–20% in COI, overlapping with that between species from different genera or families (Emerson et al., 2011; Porco et al., 2012a, 2014). COI and COII are variable both between and within species, and topologies of lineages based only on COI or COII typically were not well supported, e.g. in *Entomobrya* (Katz et al., 2015), *Orchesella* (Frati et al., 2000a), and *Protaphorura* (Sun et al., 2017). Combining more conserved nuclear markers with COI or COII proved to be an efficient way to study the phylogeography of these widely distributed Collembola species (von Saltzwedel et al., 2017, 2016). Thus in **Chapter IV**, the COII data set of the *L. lanuginosus* group, more complete than that of COI, was used to analyze genetic diversity and population structure of all lineages of the *L. lanuginosus* species group. The phylogenetic relationships of these lineages were reconstructed based on concatenated data set of COI, COII, 28S D1–2 and EF1-α, to further infer their divergence and colonization patterns in Europe.

1.2 "Cryptic" species in Collembola

Collembola are small terrestrial arthropods morphologically similar to insects, possessing three tagmata, one pair of antennae, and three pairs of thoracic legs. Collembola differ from all other hexapods by the presence of three synapomorphies: a ventral tube (or collophore), the tenaculum (or

retinaculum), and the furcula (Hopkin, 1997), but tenaculum and furcula are secondarily lost in some taxa. Traditionally, species of Collembola have been identified and described using morphological characters, i.e., following the morphological species concept (MSC) which is based on distinct discontinuity in the series of biotypes (Du Rietz, 1930). However, morphological characters may undergo convergent evolution if they are under strong functional selective pressure. The use of morphological data alone therefore may fail to identify species and this may be particularly widespread in species showing low morphological variation, i.e. comprising of cryptic species.

Accurate delimitation of species is essential for understanding species diversity and their distribution. It is generally accepted that species are independently evolving lineages and various operational criteria are used for species identification, depending on the species concept that is being invoked (De Queiroz, 2007). Among the ca. 22 species concepts (Mayden, 1997), the two most widely used species concepts are the biological species concept (BSC), which is based on reproductive isolation (Dobzhansky, 1950; Mayr, 1942), and the phylogenetic species concept (PSC), which is based on reciprocal monophyly (Baum and Shaw, 1995; Wheele and Meier, 2000). However, the BSC is rarely used in Collembola (and other arthropods), due to the limits of proving the ability to interbreed and form hybrids. Generally, however, in Collembola the MSC and BSC are compatible as indicated by a series of laboratory hybridization studies in *Hypogastruridae* (Skarzyński, 2005), and the MSC and similarity criteria are now widely applied for species diagnoses in Collembola groups (Deharveng, 2004).

There is still considerable disagreement about the hierarchy of types of characters and validation criteria for species delimitation (Padial et al., 2010; Schlick-Steiner et al., 2010), and this is particularly pronounced in Collembola. Even though a wide range of diagnostic morphological characters (chaetotaxy, mouthparts) have improved the taxonomy of Collembola (Deharveng, 2004). Considerable morphological conservatism raises difficulties in deciding which morphological trait(s) relate to species. Recent molecular studies have revealed deeply divergent lineages and high genetic diversity of morphologically conserved and widely distributed Collembola species such as *Ceratophysella denticulate* (Bagnall, 1941), *Folsomia quadrioculata* (Tullberg, 1871), *Isotomiella minor* (Schaeffer, 1896) (von Saltzwedel et al., 2016), *L. lanuginosus* (Cicconardi et al., 2010; B. Zhang et al., 2018), *Orchesella cincta* (Linnaeus, 1758) (Timmermans et al., 2005), and *Parisotoma notabilis* (Schäffer, 1896) (Porco et al., 2012b; von Saltzwedel et al., 2017). These species comprise at least two distinct lineages, and genetic distances between lineages are as high as that between species (Cicconardi et al., 2010; Porco et al., 2012a). On one hand, the cryptic species problem supports the notion that global Collembola diversity may be greatly underestimated (Cicconardi et al., 2010; Deharveng, 2004), on the other hand, it raises concerns about how to delimit cryptic species.

1.3 Molecular delimitation criteria in Collembola

Molecular data, i.e., DNA sequences, have many advantages in species delimitation. For example, data from multiple loci can be integrated, species limits can be determined without defining a *priori* species, and the statistical support for species-level decisions can be estimated (Wiens, 2007). Molecular data can provide additional information related to species identification, delimitations that integrate genetic data greatly alleviate the difficulties of using morphological taxonomy alone in Collembola (e.g., Katz et al., 2015; Sun et al., 2017; F. Zhang et al., 2018). Many DNA-based statistical methods have been developed to improve species delimitations (Knowles and Carstens, 2007; Wiens, 2007). For example the Automatic Barcode Gap Discovery (ABGD), a fast and simple distance-based method which splits sequence alignment datasets into candidate species based on a barcode gap computed from a single locus (Puillandre et al., 2012).

Other methods use single or multiple genetic markers and consider evolutionary models, which greatly improved molecular species delimitation (Hailer et al., 2012; Knowles and Carstens, 2007). For example, the Poisson Tree Processes model (PTP) tested species boundaries on non-ultrametric phylogenies by detecting significant difference in the number of substitutions between and within species (Kapli et al., 2017; Zhang et al., 2013). The Bayesian Phylogenetics and Phylogeography (BPP) program uses multilocus sequence data to delimit species, which can identify cryptic species that may be misidentified as a single species (Yang and Rannala, 2017, 2010). Since any method could possibly be violated in a particular empirical system, applying a wide range of delimitation analyses has been proposed to strengthen the confidence in the results that are congruent across methods (Carstens et al., 2013).

1.4 Phylogeny and biogeography of Collembola

Collembola, together with Protura, Diplura, and Insecta, form the four main groups of Hexapoda. Fossil record of Collembola date back to the Devonian, ca. 400 million years ago (Mya; Hirst & Maulik, 1926; Whalley & Jarzembowski, 1981). Collembola have been assigned to four orders, i.e., Poduromorpha Börner, 1913, Entomobryomorpha Börner, 1913, Symphypleona Börner, 1901 and Neelipleona Massoud, 1971, and 34 families (Bellinger et al., 1996-2018). Phylogenetic relationships of the major clades of Collembola that were reconstructed based on morphological data are incongruent with that based on molecular data, presumably due to the uncertain position of *Tomoceridae* (D'Haese, 2003, 2002; Luan et al., 2005; Schneider et al., 2011; Xiong et al., 2008; Yu et al., 2016). While, phylogenetic reconstruction based on genetic data well supported the monophyly and the relationships of the subfamilies (except Orchesellinae) of Entomobryidae, the largest family

in Collembola (Zhang et al., 2014a). Integrating molecular and traditional and new morphological evidence further revised the systematics of Entomobryidae, separating Orchesellinae into three subfamilies (Zhang and Deharveng, 2015). Collembola usually show clear specific spatial distribution patterns (Fiera and Ulrich, 2012), thus molecular phylogenetic reconstructions of Collembola at genus or species level greatly help inferring the speciation and historical biogeographical processes of Collembola (Cicconardi et al., 2013, 2010; Zhang et al., 2014b).

Collembola showed latitudinal and longitudinal gradients with species richness decreasing from south to north and from west towards east in Europe (Fiera and Ulrich, 2012; Ulrich and Fiera, 2009). Area, winter length and annual temperature difference are major predictors of species richness of European Collembola (Ulrich and Fiera, 2009). Mediterranean regions, i.e., Spain, France, and Italy are rich in endemic Collembola while higher latitudes are richer in widespread species (Deharveng et al., 2008; Fiera and Ulrich, 2012). Climatic, geographic and geologic histories are the main factors that determine the biogeography of current biota. The present-day distribution patterns of European Collembola, especially species with restricted range sizes, was suggested to be shaped by Quaternary glaciation (Fiera and Ulrich, 2012; Frenzel, 2005; Knowles, 2000), following the common pattern that species retreating to Southern refugia and re-colonizing Central Europe during interglacial periods (Hewitt, 2000, 1999; Hewitt et al., 1996). But cryptic refugia also existed in the Northern Hemisphere during glaciation (Stewart et al., 2010), and multiregional postglacial colonization, including refugia north of Alps, were suggested to contribute the current spatial distribution patterns of the European Collembola (Fiera et al., 2016; Fiera and Ulrich, 2012). Geographical isolation, constant environmental conditions, and heterogeneous geographical structures within these regions may further accelerate the local diversification (Fiera et al., 2017). Thus populations of species in Central Europe may be genetically very distinct from that from Southern Europe.

Collembola species with a wide distribution range provide good model organisms to study the effects of Quaternary glaciation on the dispersal and divergence of below ground biota in Europe. As mentioned above, these widely distributed species comprise high intra-specific variations and distinct lineages across Europe (Cicconardi et al., 2010; Porco et al., 2012b; Timmermans et al., 2005; von Saltzwedel et al., 2017, 2016). Major lineages of three Collembola species inhabiting different regions of Europe diverged during the Miocene epoch (von Saltzwedel et al., 2016), i.e., long before Quaternary glaciation, suggesting that European Collembola may not follow the common recolonization patterns (Hewitt, 2000, 1999; Hewitt et al., 1996). However, both Miocene climate change (Bruch et al., 2007) and Quaternary glaciation were proved to be the main factors that significantly affected the distribution pattern of the Antarctic Collembola (Caruso et al., 2009; McGaughran et al., 2010a, 2010b; Stevens and Hogg, 2006). Thus a study on the genetic structure of

the *L. lanuginosus* species group across Central, Southern, and Southwestern Europe will enable us to identify which and how past climate changes affected the divergence of this group of Collembola.

1.5 The European *Lepidocyrtus* species and *Lepidocyrtus lanuginosus* species group

Lepidocyrtus Bourlet, 1839 is a Collembola genus with high species richness, comprising up to 10 subgenera (Wang et al., 2003) and 225 species that are distributed all over the world (Bellinger et al., 1996-2018). The taxonomic status of *Lepidocyrtus* subgenera remains problematic due to the confusion of diagnostic morphological characters (see Mateos & Greenslade, 2015). The European *Lepidocyrtus* species were assigned to five monophyletic species groups based on both morphological and genetic data (Mateos et al., 2018). The *L. lanuginosus* species group comprises three species, i.e., *L. bicoloris* Mateos, 2012, *L. cyaneus* Tullberg, 1871, and *L. lanuginosus*, but variation in morphological characters exists within each species (Mateos, 2012). Body color is an important morphological character to distinguish the species *L. cyaneus* (entire dark blue body) and *L. lanuginosus* (entire yellow body) (Hopkin, 2007; Mateos, 2012, 2008). However, phylogenetic analysis has revealed that *L. lanuginosus* is paraphyletic with *L. cyaneus* nested within *L. lanuginosus* (Cicconardi et al., 2010; Mateos et al., 2018; B. Zhang et al., 2018). Phylogenetic reconstruction involving lineages occurring in different regions of Europe may allow evaluating if body color is a valid species marker in the *L. lanuginosus* species group.

Lepidocyrtus cyaneus commonly is considered to predominantly occur in grasslands (Auclerc et al., 2009; Migliorini et al., 2003) but also occurs in forests (Heidemann et al., 2014; Urbanovičová et al., 2014) and arable fields (Scheunemann et al., 2015); *L. lanuginosus* has been considered as habitat generalist associated with arable fields (Querner et al., 2013), grasslands (Auclerc et al., 2009; Heiniger et al., 2015), as well as forests (Cicconardi et al., 2010; Ferlian et al., 2015). However, several species of the genus *Isotoma* (Collembola, Isotomidae) show microgeographic and microhabitat segregation (Carapelli et al., 1995), indicating that closely related Collembola species differ in their preferred habitats. Considering the existence of high intra-specific genetic variations and ubiquitous distribution in variable habitats, I also investigated whether cryptic species / lineages of *L. lanuginosus* exist in different but closely connected habitats, and further tested whether these habitats function as filters sorting for specific genetic lineages.

The *L. lanuginosus* species group comprises cryptic species / lineages in the Mediterranean region and Central Europe (Cicconardi et al., 2010; B. Zhang et al., 2018). The genetic distances in COI between these lineages were considerably larger than the assumed threshold for delimiting species (Hebert et al., 2004, 2003; Hebert and Gregory, 2005) and specifically for delimiting Collembola

species (Hogg and Hebert, 2004; Sun et al., 2017; F. Zhang et al., 2018). Thus the true species richness of the *L. lanuginosus* species group likely is underestimated, and genetic based delimitation surpasses morphological delimitation in identifying their cryptic species diversity.

Both species *L. cyaneus* and *L. lanuginosus* are globally distributed and belong to the most commonly recorded epedaphic Collembola species in Europe (https://fauna-eu.org; Salmon et al., 2014). They are considered to be fast colonizers, with long legs and antenna, developed furcula and complete visual apparatus with eight ocella per eye spot (Auclerc et al., 2009; Ponge et al., 2006). Thus they are good model organisms to study their phylogeographic structure and the effects of past climate changes on the distribution and genetic structure of above-ground living Collembola.

1.6 Structure of the thesis

The aims of this thesis were to investigate the genetic structure and cryptic diversity of three Collembola species, i.e., *L. cyaneus*, *L. lanuginosus*, and *L. bicoloris* Mateos, 2012 of the *L. lanuginosus* species group at local and European scale to test if their diversity and structure are related to different habitats, to geographical regions, or historical climate changes. In chapter II, I investigate the genetic variance of specimens of *L. lanuginosus* from three dominant habitat types in Central Europe, i.e. forests, grasslands and arable fields. In Chapter III, I compare genetic distances within and between *L. cyaneus* and *L. lanuginosus* by extending the sampling regions to Central, Southern, and Southwestern Europe. In Chapter IV, I analyze lineages of these two morphospecies together to investigate the phylogeographic structure of the whole species group. Further, I investigate the genetic diversity and structure of the lineages in different regions to infer the colonization routes, and to estimate the divergence times of these lineages to gain insight into how past climate and geography may have affected the current distribution of cryptic species / lineages of the *L. lanuginosus* species group.

I examined the following main hypotheses:

(1) The species *L. lanuginosus* comprises cryptic species / lineages in a local region (within 20 km × 20 km), and these lineages are sorted by habitat types. **Chapter II**

(2) The three species in the *L. lanuginosus* species group, i.e. *L. bicoloris*, *L. lanuginosus*, and *L. cyaneus* are genetically well separated, with the genetic divergence within species being lower than between species. **Chapter III**

(3) Body color is a valid species character reflecting monophyly of each species of the *L. lanuginosus* species group. **Chapter III**

(4) Lineages of the *L. lanuginosus* species group within sampling regions are phylogenetically clustered and isolated from lineages from other sampling regions, reflecting that lineages inhabiting Central Europe originated from Southern or Southwestern Europe. **Chapter IV**

(5) Divergence times of lineages of the *L. lanuginosus* species group fall into the Quaternary suggesting that Quaternary glaciation drove the dispersal and divergence of these lineages. **Chapter IV**

References

Anslan, S., Tedersoo, L., 2015. Performance of cytochrome c oxidase subunit I (COI), ribosomal DNA Large Subunit (LSU) and Internal Transcribed Spacer 2 (ITS2) in DNA barcoding of Collembola. Eur. J. Soil Biol. 69, 1–7. doi:10.1016/j.ejsobi.2015.04.001

Auclerc, A., Ponge, J.F., Barot, S., Dubs, F., 2009. Experimental assessment of habitat preference and dispersal ability of soil springtails. Soil Biol. Biochem. 41, 1596–1604. doi:10.1016/j.soilbio.2009.04.017

Baum, D.A., Shaw, K.L., 1995. Genealogical perspectives on the species problem, in: P. C. Hoch, and A. G. Stephenson, E. (Ed.), Experimental and Molecular Approaches to Plant Biosystematics. Missouri Botanical Garden; St. Louis., pp. 289–303.

Beet, C.R., Hogg, I.D., Collins, G.E., Cowan, D.A., Wall, D.H., Adams, B.J., 2016. Genetic diversity among populations of Antarctic springtails (Collembola) within the Mackay Glacier ecotone. Genome 59, 762–770. doi:10.1139/gen-2015-0194

Bellinger, P.F., Christiansen, K.A., Janssens, F., 1996. Checklist of the Collembola of the World. http://www.collembola.org/ (accessed 6.26.18).

Bruch, A.A., Uhl, D., Mosbrugger, V., 2007. Miocene climate in Europe — Patterns and evolution. Palaeogeogr. Palaeoclimatol. Palaeoecol. 253, 1–7. doi:10.1016/j.palaeo.2007.03.030

Carapelli, A., Convey, P., Nardi, F., Frati, F., 2014. The mitochondrial genome of the antarctic springtail *Folsomotoma octooculata* (Hexapoda; Collembola), and an update on the phylogeny of collembolan lineages based on mitogenomic data. Entomologia 2. doi:10.4081/entomologia.2014.190

Carapelli, A., Frati, F., Fanciulli, P.P., Dallai, R., 1995. Genetic differentiation of six sympatric species of *Isotomurus* (Collembola, Isotomidae): is there any difference in their microhabitat preference? Eur. J. Soil Biol. 31, 87–99.

Carstens, B.C., Pelletier, T.A., Reid, N.M., Satler, J.D., 2013. How to fail at species delimitation. Mol. Ecol. 22, 4369–4383. doi:10.1111/mec.12413

Caruso, T., Hogg, I.D., Carapelli, A., Frati, F., Bargagli, R., 2009. Large-scale spatial patterns in the distribution of Collembola (Hexapoda) species in Antarctic terrestrial ecosystems. J. Biogeogr. 36, 879–886. doi:10.1111/j.1365-2699.2008.02058.x

Cicconardi, F., Fanciulli, P.P., Emerson, B.C., 2013. Collembola, the biological species concept and the underestimation of global species richness. Mol. Ecol. 22, 5382–5396. doi:10.1111/mec.12472

Cicconardi, F., Nardi, F., Emerson, B.C., Frati, F., Fanciulli, P.P., 2010. Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. Mol. Ecol. 19, 386–400. doi:10.1111/j.1365-294X.2009.04457.x

D'Haese, C.A., 2003. Morphological appraisal of Collembola phylogeny with special emphasis on Poduromorpha and a test of the aquatic origin hypothesis. Zool. Scr. 32, 563–586. doi:10.1046/j.1463-6409.2003.00134.x

D'Haese, C.A., 2002. Were the first springtails semi-aquatic? A phylogenetic approach by means of 28S rDNA and optimization alignment. Proc. R. Soc. B Biol. Sci. 269, 1143–1151. doi:10.1098/rspb.2002.1981

De Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886. doi:10.1080/10635150701701083

Deharveng, L., 2004. Recent advances in Collembola systematics. Pedobiologia. 48, 415–433. doi:10.1016/j.pedobi.2004.08.001

Deharveng, L., D'Haese, C.A., Bedos, A., 2008. Global diversity of springtails (Collembola; Hexapoda) in freshwater. Hydrobiologia 595, 329–338. doi:10.1007/s10750-007-9116-z

Dell'Ampio, E., Carapelli, A., Frati, F., 2002. Secondary structure and sequence variation of the 28S rRNA gene in the Neanuridae, and its utility as a phylogenetic marker. Pedobiologia. 46, 274–283. doi:10.1016/S0031-4056(04)70143-2

Ding, Y., Yu, D., Guo, W., Li, J., Zhang, F., 2018. Molecular phylogeny of *Entomobrya* (Collembola: Entomobryidae) from China: Color pattern groups and multiple origins. Insect Sci. 1–11. doi:10.1111/1744-7917.12559

Dobzhansky, T., 1950. Mendelian populations and their evolution. Am. Nat. 84, 401–418. doi:10.1086/281638

Du Rietz, G.E., 1930. The fundamental units of biological taxonomy. Sven. Bot. Tidskr. 24, 333–428.

Emerson, B.C., Cicconardi, F., Fanciulli, P.P., Shaw, P.J.A., 2011. Phylogeny, phylogeography, phylobetadiversity and the molecular analysis of biological communities. Philos. Trans. R. Soc. B Biol. Sci. 366, 2391–2402. doi:10.1098/rstb.2011.0057

Ferlian, O., Klarner, B., Langeneckert, A.E., Scheu, S., 2015. Trophic niche differentiation and utilisation of food resources in collembolans based on complementary analyses of fatty acids and stable isotopes. Soil Biol. Biochem. 82, 28–35. doi:10.1016/j.soilbio.2014.12.012

Fiera, C., Habel, J.C., Kunz, M., Ulrich, W., 2017. Large-scale phylogenetic structure of European springtails (Collembola) depends on species range size and postglacial colonization history. Biol. J. Linn. Soc. 120, 664–674. doi:10.1111/bij.12907

Fiera, C., Habel, J.C., Kunz, M., Ulrich, W., 2016. Large-scale phylogenetic structure of European springtails (Collembola) depends on species range size and postglacial colonization history. Biol. J. Linn. Soc. doi:10.1111/bij.12907

Fiera, C., Ulrich, W., 2012. Spatial patterns in the distribution of European springtails (Hexapoda: Collembola). Biol. J. Linn. Soc. 105, 498–506. doi:10.1111/j.1095-8312.2011.01816.x

Frati, F., Carapelli, A., 1999. An assessment of the value of nuclear and mitochondrial genes in elucidating the origin and evolution of *Isotoma klovstadi* Carpenter (Insecta, Collembola). Antarct. Sci. 11, 160–174. doi:10.1017/S0954102099000231

Frati, F., Dell'Ampio, E., 2000. Molecular phylogeny of three subfamilies of the Neanuridae (Insecta, Collembola) and the position of the antarctic species *Friesea grisea* Schäffer. Pedobiologia. 44, 342–360. doi:10.1078/S0031-4056(04)70054-2

Frati, F., Dell'Ampio, E., Casasanta, S., Carapelli, A., Paolo Fanciulli, P., 2000a. Large amounts of genetic divergence among Italian species of the genus *Orchesella* (Insecta, Collembola) and the relationships of two new species. Mol. Phylogenet. Evol. 17, 456–461. doi:10.1006/mpev.2000.0854

Frati, F., Fanciulli, P.P., Carapelli, A., Dell'Ampio, E., Nardi, F., Spinsanti, G., Dallai, R., 2000b. DNA sequence analysis to study the evolution of Antarctic Collembola. Ital. J. Zool. 67, 133–139. doi:10.1080/11250000009356368

Frati, F., Simon, C., Sullivan, J., Swofford, D.L., 1997. Evolution of the mitochondrial cytochrome oxidase II gene in Collembola. J. Mol. Evol. 44, 145–158. doi:10.1007/PL00006131

Frenzel, B., 2005. History of flora and vegetation during the Quaternary North America, in: Progress in Botany. Springer-Verlag, Berlin/Heidelberg, pp. 409–440. doi:10.1007/3-540-27043-4_17

Hailer, F., Kutschera, V.E., Hallström, B.M., Klassert, D., Fain, S.R., Leonard, J.A., Arnason, U., Janke, A., 2012. Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. Science (80-.). 336, 344–347. doi:10.1126/science.1216424

Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B Biol. Sci. 270, 313–321. doi:10.1098/rspb.2002.2218

Hebert, P.D.N., Gregory, T.R., 2005. The Promise of DNA Barcoding for Taxonomy. Syst. Biol. 54, 852–859. doi:10.1080/10635150500354886

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., Hallwachs, W., 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Natl. Acad. Sci. 101, 14812–14817. doi:10.1073/pnas.0406166101

Heidemann, K., Hennies, A., Schakowske, J., Blumenberg, L., Ruess, L., Scheu, S., Maraun, M., 2014. Free-living nematodes as prey for higher trophic levels of forest soil food webs. Oikos 123, 1199–1211. doi:10.1111/j.1600-0706.2013.00872.x

Heiniger, C., Barot, S., Ponge, J.-F., Salmon, S., Meriguet, J., Carmignac, D., Suillerot, M., Dubs, F., 2015. Collembolan preferences for soil and microclimate in forest and pasture communities. Soil Biol. Biochem. 86, 181–192. doi:10.1016/j.soilbio.2015.04.003

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. Nature 405, 907–913. doi:10.1038/35016000

Hewitt, G.M., 1999. Post-glacial re-colonization of European biota. Biol. J. Linn. Soc. 68, 87–112. doi:10.1006/bijl.1999.0332

Hewitt, G.M., Prendini, L., Savary, W.E., et al., 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biol. J. Linn. Soc. 58, 247–276. doi:10.1111/j.1095-8312.1996.tb01434.x

Hirst, S., Maulik, S., 1926. On some arthropod remains from the Rhynie chert (Old Red Sandstone). Geol. Mag. 63, 69–71. doi:10.1017/S0016756800083692

Hogg, I.D., Hebert, P.D.., 2004. Biological identification of springtails (Hexapoda: Collembola) from the Canadian Arctic, using mitochondrial DNA barcodes. Can. J. Zool. 82, 749–754. doi:10.1139/z04-041

Hopkin, S.P., 2007. A Key to the Springtails (Collembola) of Britain and Ireland, Field Studies Council (AIDGAP Project). FSC Publicatio, Shrewsbury.

Hopkin, S.P., 1997. Biology of the springtails: (Insecta: Collembola). Oxford University Press, Oxford, UK.

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. doi:10.1093/bioinformatics/btx025

Katz, A.D., Giordano, R., Soto-Adames, F.N., 2015. Operational criteria for cryptic species delimitation when evidence is limited, as exemplified by North American *Entomobrya* (Collembola: Entomobryidae). Zool. J. Linn. Soc. 173, 818–840. doi:10.1111/zoj.12220

Knowles, L.L., 2000. Tests of pleistocene speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of western North America. Evolution (N. Y). 54, 1337–1348. doi:10.1111/j.0014-3820.2000.tb00566.x

Knowles, L.L., Carstens, B.C., 2007. Delimiting species without monophyletic gene trees. Syst. Biol. 56, 887–895. doi:10.1080/10635150701701091

Luan, Y.X., Mallatt, J.M., Xie, R.D., Yang, Y.M., Yin, W.Y., 2005. The phylogenetic positions of three basal-hexapod groups (Protura, Diplura, and Collembola) based on ribosomal RNA gene sequences. Mol. Biol. Evol. 22, 1579–1592. doi:10.1093/molbev/msi148

Mateos, E., 2012. The European *Lepidocyrtus lanuginosus* group (Collembola: Entomobryidae), definition and description of a new species from Spain. Zootaxa 3570, 69–81.

Mateos, E., 2008. The European *Lepidocyrtus* Bourlet, 1839 (Collembola: Entomobryidae). Zootaxa 1769, 35–59.

Mateos, E., Escuer, P., Buşmachiu, G., Riutort, M., Álvarez-Presas, M., 2018. Untangling *Lepidocyrtus* (Collembola, Entomobryidae): new molecular data shed light on the relationships of the European groups. Invertebr. Syst. 32, 639–651. doi:10.1071/IS17056

Mateos, E., Greenslade, P., 2015. Towards understanding *Lepidocyrtus* Bourlet, 1839 (Collembola, Entomobryidae) I: Diagnosis of the subgenus *Setogaster*, new records and redescriptions of species. Zootaxa 4044, 105–129. doi:10.11646/zootaxa.4044.1.6

Mayden, R.L., 1997. A hierarchy of species concepts: the denouement in the saga of the species problem, in: M. F. Claridge, H. A. Dawah, M.R.W. (Ed.), Species: The Units of Diversity. Chapman and Hall, London, pp. 381–423.

Mayr, E., 1942. Systematics and the origin of species from the viewpoint of a zoologist. Columbia University Press, New York.

McGaughran, A., Stevens, M.I., Holland, B.R., 2010a. Biogeography of circum-Antarctic springtails. Mol. Phylogenet. Evol. 57, 48–58. doi:10.1016/j.ympev.2010.06.003

McGaughran, A., Torricelli, G., Carapelli, A., Frati, F., Stevens, M.I., Convey, P., Hogg, I.D., 2010b. Contrasting phylogeographical patterns for springtails reflect different evolutionary histories between the Antarctic Peninsula and continental Antarctica. J. Biogeogr. 37, 103–119. doi:10.1111/j.1365-2699.2009.02178.x

Migliorini, M., Fanciulli, P.P., Bernini, F., 2003. Comparative analysis of two edaphic zoocoenoses (Acari Oribatida; Hexapoda Collembola) in the area of Orio al Serio Airport (Bergamo, northern Italy). Pedobiologia. 47, 9–18. doi:10.1078/0031-4056-00164

Padial, J.M., Miralles, A., De la Riva, I., Vences, M., 2010. The integrative future of taxonomy. Front. Zool. 7, 16. doi:10.1186/1742-9994-7-16

Ponge, J.F., Dubs, F., Gillet, S., Sousa, J.P., Lavelle, P., 2006. Decreased biodiversity in soil springtail communities: the importance of dispersal and landuse history in heterogeneous landscapes. Soil Biol. Biochem. 38, 1158–1161. doi:10.1016/j.soilbio.2005.09.004

Porco, D., Bedos, A., Greenslade, P., Janion, C., Skarzyński, D., Stevens, M.I., Jansen Van Vuuren, B., Deharveng, L., 2012a. Challenging species delimitation in Collembola: cryptic diversity among common springtails unveiled by DNA barcoding. Invertebr. Syst. 26, 470–477. doi:10.1071/IS12026

Porco, D., Potapov, M., Bedos, A., Busmachiu, G., Weiner, W.M., Hamra-Kroua, S., Deharveng, L., 2012b. Cryptic diversity in the ubiquist species *Parisotoma notabilis* (Collembola, Isotomidae): a long-used chimeric species? PLoS One 7, e46056. doi:10.1371/journal.pone.0046056

Porco, D., Skarzyński, D., Decaëns, T., Hebert, P.D.N., Deharveng, L., 2014. Barcoding the Collembola of Churchill: A molecular taxonomic reassessment of species diversity in a sub-Arctic area. Mol. Ecol. Resour. 14, 249–261. doi:10.1111/1755-0998.12172

Puillandre, N., Lambert, A., Brouillet, S., Achaz, G., 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol. Ecol. 21, 1864–1877. doi:10.1111/j.1365-294X.2011.05239.x

Querner, P., Bruckner, A., Drapela, T., Moser, D., Zaller, J.G., Frank, T., 2013. Landscape and site effects on Collembola diversity and abundance in winter oilseed rape fields in eastern Austria. Agric. Ecosyst. Environ. 164, 145–154. doi:10.1016/j.agee.2012.09.016

Salmon, S., Ponge, J.F., Gachet, S., Deharveng, L., Lefebvre, N., Delabrosse, F., 2014. Linking species, traits and habitat characteristics of Collembola at European scale. Soil Biol. Biochem. 75, 73–85. doi:10.1016/j.soilbio.2014.04.002

Scheunemann, N., Digel, C., Scheu, S., Butenschoen, O., 2015. Roots rather than shoot residues drive soil arthropod communities of arable fields. Oecologia 179, 1135–1145. doi:10.1007/s00442-015-3415-2

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. Annu. Rev. Entomol. 55, 421–438. doi:10.1146/annurev-ento-112408-085432

Schneider, C., Cruaud, C., D'Haese, C. a., 2011. Unexpected diversity in Neelipleona revealed by molecular phylogeny approach (Hexapoda, Collembola). Soil Org. 83, 383–398.

Schneider, C., D'Haese, C.A., 2013. Morphological and molecular insights on *Megalothorax*: The largest Neelipleona genus revisited (Collembola). Invertebr. Syst. 27, 317–364. doi:10.1071/IS13002

Skarzyński, D., 2005. Taxonomy of the ceratophysellan lineage (Collembola: Hypogastruridae) in the light of laboratory hybridization studies. Zootaxa 1043, 47–59. doi:10.11646/zootaxa.1043.1.4

Soto-Adames, F.N., 2000. Phylogeny of neotropical *Lepidocyrtus* (Collembola: Entomobryidae): First assessment of patterns of speciation in Puerto Rico and phylogenetic relevance of some subgeneric diagnostic characters. Syst. Entomol. 25, 485–502. doi:10.1046/j.1365-3113.2000.00127.x

Stevens, M.I., Frati, F., McGaughran, A., Spinsanti, G., Hogg, I.D., 2007. Phylogeographic structure suggests multiple glacial refugia in northern Victoria Land for the endemic Antarctic springtail *Desoria klovstadi* (Collembola, Isotomidae). Zool. Scr. 36, 201–212. doi:10.1111/j.1463-6409.2006.00271.x

Stevens, M.I., Hogg, I.D., 2006. Contrasting levels of mitochondrial DNA variability between mites (Penthalodidae) and springtails (Hypogastruridae) from the Trans-Antarctic Mountains suggest long-term effects of glaciation and life history on substitution rates, and speciation processes. Soil Biol. Biochem. 38, 3171–3180. doi:10.1016/j.soilbio.2006.01.009

Stewart, J.R., Lister, A.M., Barnes, I., Dalen, L., 2010. Refugia revisited: individualistic responses of species in space and time. Proc. R. Soc. B Biol. Sci. 277, 661–671. doi:10.1098/rspb.2009.1272

Sun, X., Zhang, F., Ding, Y., Davies, T.W., Li, Y., Wu, D., 2017. Delimiting species of *Protaphorura* (Collembola: Onychiuridae): integrative evidence based on morphology, DNA sequences and geography. Sci. Rep. 7, 8261. doi:10.1038/s41598-017-08381-4

Timmermans, M.J.T.N., Ellers, J., Mariën, J., Verhoef, S.C., Ferwerda, E.B., Van Straalen, N.M., 2005. Genetic structure in *Orchesella cincta* (Collembola): Strong subdivision of European populations inferred from mtDNA and AFLP markers. Mol. Ecol. 14, 2017–2024. doi:10.1111/j.1365-294X.2005.02548.x

Ulrich, W., Fiera, C., 2009. Environmental correlates of species richness of European springtails (Hexapoda: Collembola). Acta Oecologica 35, 45–52. doi:10.1016/j.actao.2008.07.007

Urbanovičová, V., Miklisová, D., Kováč, Ľ., 2014. Forest disturbance enhanced the activity of epedaphic Collembola in windthrown stands of the High Tatra mountains. J. Mt. Sci. 11, 449–463. doi:10.1007/s11629-013-2736-z

von Saltzwedel, H., Scheu, S., Schaefer, I., 2017. Genetic structure and distribution of *Parisotoma notabilis* (Collembola) in Europe: Cryptic diversity, split of lineages and colonization patterns. PLoS One 12, e0170909. doi:10.1371/journal.pone.0170909

von Saltzwedel, H., Scheu, S., Schaefer, I., 2016. Founder events and pre-glacial divergences shape the genetic structure of European Collembola species. BMC Evol. Biol. 16, 148. doi:10.1186/s12862-016-0719-8

Wang, F., Chen, J.X., Christiansen, K., 2003. Taxonomy of the genus *Lepidocyrtus* s.l. (Collembola: Entomobryidae) in East and Southeast Asia and Malaysia, with description of a new species from the People's Republic of China. Can. Entomol. 135, 823–837. doi:10.4039/n02-106

Whalley, P., Jarzembowski, E.A., 1981. A new assessment of Rhyniella, the earliest known insect, from the Devonian of Rhynie, Scotland. Nature 291, 317. doi:10.1038/291317a0

Wheele, Q., Meier, R., 2000. Species concepts and phylogenetic theory: a debate. Columbia University Press, New York.

Wiens, J.J., 2007. Species delimitation: new approaches for discovering diversity. Syst. Biol. 56, 875–878. doi:10.1080/10635150701748506

Xiong, Y., Gao, Y., Yin, W.Y., Luan, Y.X., 2008. Molecular phylogeny of Collembola inferred from ribosomal RNA genes. Mol. Phylogenet. Evol. 49, 728–735. doi:10.1016/j.ympev.2008.09.007

Yang, Z., Rannala, B., 2017. Bayesian species identification under the multispecies coalescent provides significant improvements to DNA barcoding analyses. Mol. Ecol. 26, 3028–3036. doi:10.1111/mec.14093

Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. 107, 9264–9269. doi:10.1073/pnas.0913022107

Yu, D., Ding, Y., Ma, Y., 2017. Revision of *Tomocerus similis* Chen & Ma, with discussion of the *kinoshitai* complex and the distal tibiotarsal chaetae in Tomocerinae (Collembola, Tomoceridae). Zootaxa 4268, 395–410. doi:10.11646/zootaxa.4268.3.5

Yu, D., Zhang, F., Stevens, M.I., Yan, Q., Liu, M., Hu, F., 2016. New insight into the systematics of Tomoceridae (Hexapoda, Collembola) by integrating molecular and morphological evidence. Zool. Scr. 45, 286–299. doi:10.1111/zsc.12149

Zhang, B., Chen, T.W., Mateos, E., Scheu, S., Schaefer, I., 2018. Cryptic species in *Lepidocyrtus lanuginosus* (Collembola: Entomobryidae) are sorted by habitat type. Pedobiologia. 68, 12–19. doi:10.1016/j.pedobi.2018.03.001

Zhang, F., Chen, Z., Dong, R., Deharveng, L., Stevens, M.I., Huang, Y., Zhu, C., 2014a. Molecular phylogeny reveals independent origins of body scales in Entomobryidae (Hexapoda: Collembola). Mol. Phylogenet. Evol. 70, 231–239. doi:10.1016/j.ympev.2013.09.024

Zhang, F., Deharveng, L., 2015. Systematic revision of Entomobryidae (Collembola) by integrating molecular and new morphological evidence. Zool. Scr. 44, 298–311. doi:10.1111/zsc.12100

Zhang, F., Jantarit, S., Nilsai, A., Stevens, M.I., Ding, Y., Satasook, C., 2018. Species delimitation in the morphologically conserved *Coecobrya* (Collembola: Entomobryidae): A case study integrating morphology and molecular traits to advance current taxonomy. Zool. Scr. 47, 342–356. doi:10.1111/zsc.12279

Zhang, F., Sun, D.D., Yu, D.Y., Wang, B.X., 2015. Molecular phylogeny supports S-chaetae as a key character better than jumping organs and body scales in classification of Entomobryoidea (Collembola). Sci. Rep. 5, 12471. doi:10.1038/srep12471

Zhang, F., Yu, D., Luo, Y., Ho, S.Y.W., Wang, B., Zhu, C., 2014b. Cryptic diversity, diversification and vicariance in two species complexes of *Tomocerus* (Collembola, Tomoceridae) from China. Zool. Scr. 43, 393–404. doi:10.1111/zsc.12056

Zhang, J., Kapli, P., Pavlidis, P., Stamatakis, A., 2013. A general species delimitation method with applications to phylogenetic placements. Bioinformatics 29, 2869–2876. doi:10.1093/bioinformatics/btt499

Chapter II

Cryptic species in *Lepidocyrtus lanuginosus* (Collembola: Entomobryidae) are sorted by habitat type

Bing Zhang, Ting-Wen Chen, Eduardo Mateos, Stefan Scheu, Ina Schaefer

Abstract

High intraspecific genetic variance in Collembola indicates that cryptic species are widespread and this challenges the delimitation of morphologically defined species. *Lepidocyrtus lanuginosus* (Gmelin, 1788) is a widely distributed habitat generalist with high genetic variance between populations from different locations in Europe. In this study we investigated the genetic variance of *L. lanuginosus* from three dominant habitat types in Central Europe, i.e. forests, grasslands and arable fields, using four molecular markers (ribosomal subunit 28S rDNA D1–2 domain, elongation factor 1-α, cytochrome c oxidase subunit I and subunit II). The results suggest that *L. lanuginosus* separates into three major genetic lineages with one of these lineages being close to *Lepidocyrtus cyaneus* Tullberg, 1871. The phylogenetic tree based on the concatenated data set of four genes suggests that all lineages of *L. lanuginosus* are monophyletic. Selective colonization of the three habitats by these lineages indicates that they are sorted by habitats: one lineage was common and occurred in each of the three habitat types but preferentially in arable land, the second was restricted to forest, and the third, although rare, preferentially occurred in grassland. Our results indicate that genetic markers allow delineating cryptic species in Collembola, which are widespread, morphologically coherent and differ in the habitats they colonize. The existence of cryptic species / lineages in widely distributed Collembola species that sort by habitat type calls for studies integrating genetic structure and ecological traits.

Keywords: DNA barcoding; genetic variation; *Lepidocyrtus cyaneus*; mitochondrial genes; pigmentation; springtail

2.1 Introduction

The development of molecular methods in the past decade triggered the discovery of cryptic species (Bickford et al., 2007; Emerson et al., 2011), in particular among soil organisms. Soil animal species show exceptionally high genetic variability as documented for earthworms (James et al., 2010; King et al., 2008), oribatid mites (Rosenberger et al., 2013; Schäffer et al., 2010) and springtails (Cicconardi et al., 2013, 2010). High genetic variability of morphospecies suggests the existence of cryptic species, which may substantially contribute to the diversity of belowground invertebrates (Porco et al., 2012a). Cryptic species are morphologically indistinguishable but show genetic differentiation, suggesting that they differ in their ecological niches (Bidochka et al., 2001; Davidson-Watts et al., 2006) as a result of niche-differentiation within a habitat or adaptation to different habitats (Eisenring et al., 2016; Tarjuelo et al., 2001). Environmental differences can impede gene flow

across habitat boarders, promoting population divergence (isolation by environment; Wang and Summers, 2010). Springtails (Hexapoda: Collembola) are among the most abundant and diverse soil invertebrates and occur in virtually any terrestrial habitat (Hopkin, 1997). Currently, about 8,600 species have been described. Traditionally, delimitation of Collembola species relies on morphological characters, mainly chaetotaxy, a method that examines arrangement of chaetae on different body parts (Katz et al., 2015). Molecular studies, however, indicated that the number of existing Collembola species likely is much higher as many species comprise a number of cryptic species (Emerson et al., 2011; Porco et al., 2012a; Soto-Adames, 2002). Notably, cryptic species appear to be common in widely distributed and locally abundant Collembola species, such as *Orchesella cincta* (Linnaeus, 1758) (Timmermans et al., 2005) and *Parisotoma notabilis* (Schäffer, 1896) (Porco et al., 2012b; von Saltzwedel et al., 2017).

Lepidocyrtus lanuginosus (Gmelin, 1788) is one of the most widely distributed Collembola species in Central and Western Europe (Salmon et al., 2014). It rapidly colonizes new habitats and is characterized by long legs and antenna, developed furcula and complete visual apparatus with eight ocella per eye spot (Ponge et al., 2006). *Lepidocyrtus lanuginosus* has been considered as habitat generalist (Auclerc et al., 2009) associated with arable fields (Querner et al., 2013) and grasslands (Auclerc et al., 2009; Heiniger et al., 2015), i.e. anthropogenic habitats characterized by disturbances of varying intensity. However, the species also colonizes forests (Cicconardi et al., 2010), which are rather stable habitats. In forests of the Mediterranean region, *L. lanuginosus* displays high genetic variation, suggesting the existence of cryptic species or impeded gene flow between forests (Cicconardi et al., 2010). It is therefore likely that *L. lanuginosus* of different habitats also comprises cryptic species or lineages with distinct genetic structure.

In this study we investigated phylogenetic relationships and genetic distances between populations of *L. lanuginosus* from three different habitats in Central Europe, i.e. arable fields, grasslands and forests that were replicated at six locations. Geographic distances between sampling locations were considerably larger than between habitat types allowing to identify habitat specific genotypes. Intraspecific genetic variation was measured using two mitochondrial (COI and COII) and two nuclear genes (28S ribosomal DNA D1–2 domain and elongation factor 1-α). Mitochondrial COI is commonly used for animal DNA barcoding and reliably distinguishes between Collembola species as well as cryptic species (Hebert et al., 2003; Hogg and Hebert, 2004; Porco et al., 2012a), while COII has been used to reconstruct phylogenetic relationships of Collembola including the genus *Lepidocyrtus* (Cicconardi et al., 2013, 2010; Frati et al., 2000; Stevens et al., 2007). The two nuclear genes represent markers independent from mitochondrial genes strengthening the delineation of phylogenetic relationships among individuals of each lineage of *L. lanuginosus*. We checked if the phylogenetic

structure of *L. lanuginosus* is best explained by habitat types or by sampling locations, i.e. geographic distance. If *L. lanuginosus* comprises a single generalist species individuals should sort by geographic distances due to limited dispersal, or generate a random pattern of relatedness in the phylogenetic tree if locations and habitats were colonized multiple times. In contrast, if *L. lanuginosus* comprises cryptic species that colonize certain habitat types, genetic lineages should sort by habitat types rather than locations. Further, we compared intraspecific genetic variability of *L. lanuginosus* with genetic distances between *L. lanuginosus* and five other species of the same genus. These species included *Lepidocyrtus cyaneus* Tullberg, 1871 which is morphologically very similar to *L. lanuginosus* and belongs to the *L. lanuginosus* group (Mateos, 2012), but it differs from *L. lanuginosus* by its purple body color (Fjellberg, 2007; Hopkin, 2007; Mateos, 2008).

2.2 Material and methods

2.2.1 Sampling of animals and determination

The Collembola species *L. lanuginosus* and *L. cyaneus* were collected from six locations near the city of Göttingen, Germany; distances between sampling locations were between 4 and 28 km (supplementary Figure S2.1). Each sampling location encompassed three types of habitats: arable field, grassland and forest that were close to each other with distances below 1 km. Specimens in grasslands and arable fields were collected using an aspirator (diameter of opening 14 cm). In forests Collembola were collected from litter and extracted by heat (Kempson et al., 1963). The collected litter from forest soil covered an area that was similar to the aspirator opening used in grasslands and arable fields. Animals were preserved in 96% EtOH and stored at -20°C until further analyses. Collembola were sorted using a dissecting microscope and identified using Hopkin (2007). In total 58 individuals of *L. lanuginosus*, six individuals of *L. cyaneus* and five to six individuals of four other *Lepidocyrtus* species (*L.* cf. *arrabonicus* Traser, 2000, *L. paradoxus* Uzel, 1890, *L.* cf. *violaceus* Lubbock, 1873, and *L.* cf. *weidneri* Hüther, 1971) were analyzed. The sites at which the analyzed specimens of *L. lanuginosus* were sampled are given in Table 2.1. Prior to DNA extraction, we cut off the head of each specimen and only placed thorax and abdomen into the extraction solution. After the DNA extraction, the cuticle of the thorax and abdomen was rescued, rinsed with 96% EtOH and mounted on slides together with the head of the respective individual for inspection of the cephalic and dorsal chaetotaxy under a phase contrast microscope (Zeiss Axio Scope A1, Jena, Germany) using Mateos (2008). Microscopic slides are kept in the collection of the J.F. Blumenbach Institute of Zoology and Anthropology, University of Göttingen, Germany.

Table 2.1 Numbers of individuals of *Lepidocyrtus lanuginosus* analyzed from three habitats at the six sampling sites. The lineage that the sampled specimens were ascribed to is given in brackets. Names of sampling sites follow nearby villages; for exact geographic locations see supplementary Figure S2.1.

Site No.	Sampling site	Arable land	Grassland	Forest
1	Herberhausen	4 (L1)	5 (L3)	5 (L2)
2	Deppoldshausen	5 (L1)	0	5 (L2)
3	Ellershausen	3 (L1)	0	5 (L2)
4	Ossenfeld	1 (L1)	0	3 (L2)
5	Waake	2 (L1) / 3 (L3)	0	5 (L2)
6	Billingshausen	3 (L1)	4 (L1)	5 (L1)

2.2.2 DNA extraction and PCR

DNA of the thorax and abdomen of single individuals was extracted using the DNeasy® Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. Purified DNA was eluted in 30 µl HPLC water for PCR. PCRs of the nuclear markers elongation factor 1-α (EF1-α) and 28S rDNA (D1–2 region), and the mitochondrial markers COI and COII were performed separately in 25 µl volumes containing 12.5 µl SuperHot Taq Mastermix (Genaxxon Bioscience GmbH, Ulm, Germany) with 1.5 µl of each primer (10 pM), 3 µl H_2O, 1.5 µl $MgCl_2$ (25 mM) and 5 µl template DNA. A 208 bp fragment of the nuclear EF1-α exon was amplified using the primers EFLcuJ: 5'-ATG GGG GCA AGA TAG CGT CAA-3' and EFLcuN: 5'-TGA AGG CTG AAC GTG AAC GTG G-3' (Cicconardi et al., 2010). The ~760 bp fragment domain of D1 and D2 loop of the nuclear 28S rDNA was amplified using the primers C1': 5'-ACC CGC TGA ATT TAA GCA T-3'and D2coll: 5'- ACC ACG CAT GCW TTA GAT TG -3' (D'Haese, 2002). A 709 bp fragment of the mitochondrial COI gene was amplified using four primers LCO1490: 5'-GGT CAA CAA ATC ATA AAG ATA TTG G-3' and HCO2198: 5'-TAA ACT TCA GGG TGA CCA AAA AAT CA-3' (Folmer et al., 1994); ColFol-for: 5'-TTT CAA CAA ATC ATA ARG AYA TYG G-3' and ColFol-rev: 5'-TAA ACT TCN GGR TGN CCA AAA AAT CA-3' (Ramirez-Gonzalez et al., 2013). A 681 bp fragment of the mitochondrial COII gene was amplified using primers tRNA-K-LcuJ: 5'-GAG CGT ATT ATA AAG CGG TTT AAG-3' and tRNA-L-LcuN: 5'-CAG ACT AGT GCC ATG AAT TTA AGC-3' (Cicconardi et al., 2010). PCR conditions included one initial activation step at 95˚C for 15 min, followed by 35 amplification cycles of denaturation at 94˚C (COI and 28S D1–2) or 95˚C (COII and EF1-α) for 30 s, annealing at 45˚C (COI) or 56˚C (COII and EF1-α) for 30 s or 50˚C (28S D1–2) for 45 s, and elongation at 72˚C for 30 s, and a final elongation step at 72˚C for 6 min. PCR products were purified with Genaxxon PCR Purification Kit (Genaxxon Bioscience GmbH, Ulm, Germany) following the manufacturer's protocol, eluted in 30 µl HPLC water and sent for sequencing in the Göttingen Genome Laboratory (Institute for Microbiology

and Genetics, University of Göttingen). All sequences are available at NCBI GenBank (MH143920-MH144050, MH160100-MH160185, and MH177993-MH178033, supplementary Table S2.1).

2.2.3 Data analyses

Sequences were checked by eye aided by the chromatogras and ambiguous positions were corrected by hand. Nucleotide sequences were aligned separately for each marker using ClustalW (Thompson et al., 1994) implemented in BioEdit v7.2.5 (Hall, 1999). To analyze relationships and identify lineages among all 58 individuals of *L. lanuginosus*, phylogenetic trees were constructed based on 28S D1–2 and COII with *Orchesella villosa* (Geoffroy, 1764) as outgroup. The best fit model of sequence evolution for each marker was estimated by jModelTest v2.1.4 (Posada, 2008). Phylogenetic trees were inferred using Bayesian Inference in MrBayes v3.2 (Ronquist et al., 2012). For Bayesian Inference lset parameters were nst = 2 and rates = gamma for 28S D1–2 and nst = 2 and rates = propinv for COII, the MCMC (Markov Chain Monte Carlo) chains were run for six million generations that were sampled every 6,000th generation. For the 1,000 sampled generations a burnin of 250 was used, eliminating the first 25% of the remaining generations. Intra- and inter-lineage and intra- and inter-specific sequence divergences of 28S D1–2, EF1-α, COI and COII were calculated based on Kimura two-parameter distances (K2P) and uncorrected p-distances using Mega 5.1 (Tamura et al., 2011). Individuals with identical 28S D1–2 sequences and less than 4% K2P distances of COII were defined as one lineage (Porco et al., 2012b).

We used six individuals from each lineage of *L. lanuginosus* and *L. cyaneus* and one individual of the extra four species from genus *Lepidocyrtus* to construct a phylogenetic tree. *Orchesella villosa* was chosen as outgroup. In a 2,282 bp concatenated alignment four genes were merged: 28S D1–2, COII, COI and EF1-α of a length of 740 bp, 681 bp, 651 bp and 210 bp, respectively. The partitioned dataset was analyzed using MrBayes v3.2.6 on online CIPRES services (Miller et al., 2010). All three coding positions of the protein-coding genes COI, COII and EF1-α were included in the analyses. Best-fitting substitution models were assessed for each locus (partition) under the BIC criterion in PartitionFinder 2.1.1 (Lanfear et al., 2012).

Habitat preference of each lineage of *L. lanuginosus* was characterized using the IndVal index (Dufrêne and Legendre, 1997) which combines the specificity of a lineage for a habitat type (maximized when the lineage is only present in the habitat type analyzed) and its fidelity to this habitat (maximized when the lineage is present in all samples of the habitat type analyzed): $Ind_{ij} = A_{ij} \times B_{ij} \times 100$, with A_{ij}, the average abundance of lineage *i* in samples of habitat *j* divided by the sum of the average abundance of lineage *i* in all habitats, and B_{ij}, the number of samples of habitat *j* where the

lineage was present divided by the total number of samples of habitat *j*. Ind_{ij} ranges from 0, when lineage *i* is absent from habitat *j*, to 100, when lineage *i* is present in all samples of habitat *j* and absent in all other habitat samples. We obtained three IndVal values for each lineage of *L. lanuginosus*, i.e., IndA, IndG and IndF indicating habitat preference for arable land, grassland and forest, respectively.

To test whether there is a significant habitat association of each lineage of *L. lanuginosus,* we simulated the habitat distributions of the 58 examined individuals by randomizing their habitats 9,999 times. The observed number of a lineage in a habitat was compared to that derived from the simulations and a p-value was calculated based on the rank of the observed number as compared to the simulated ones. A p-value smaller than 0.05 indicated significant association (or avoidance) of the respective habitat by the lineage.

2.3 Results

2.3.1 Habitat sorting of lineages of *Lepidocyrtus lanuginosus*

All individuals studied under phase contrast microscope shared the characters of *L. lanuginosus* and *L. cyaneus* as indicated in Mateos (2008). Genetic analyses based on the nuclear marker 28S D1–2 and the mitochondrial marker COII both revealed three distinct lineages of *L. lanuginosus,* referred to as *L. lanuginosus* lineage 1, 2 and 3 (L1, L2, L3) (Figure 2.1). Among the 58 sequenced specimens, 21 individuals were from arable fields, 9 from grasslands and 28 from forests. Most individuals (n=27) belonged to *L. lanuginosus* L1, which occurred in each of the three habitat types but was significantly associated with arable land (IndA=66.7, IndG = 2.5, IndF = 3.1; Figure 2.2). *Lepidocyrtus lanuginosus* L2 also was common (n=23), but was significantly associated with forest (IndA = 0, IndG = 0, IndF=83.3). *Lepidocyrtus lanuginosus* L3 was rare (n=8), but was significantly associated with grassland (IndA = 6.3, IndG=10.4, IndF = 0).

At least two *L. lanuginosus* lineages were present in each habitat (Figure 2.1 and Table 2.1). Arable fields were dominated by *L. lanuginosus* L1, while *L. lanuginosus* L3 only occurred in one of the six arable fields. In grassland *L. lanuginosus* L1 and L3 were present but both were rare; each lineage occurred in only one location. In forest, *L. lanuginosus* L2 dominated and *L. lanuginosus* L1 occurred in only one of the forest locations.

2.3.2 Genetic distances and variances and phylogenetic analysis

In each of the four markers, K2P and uncorrected p-distances between lineages were

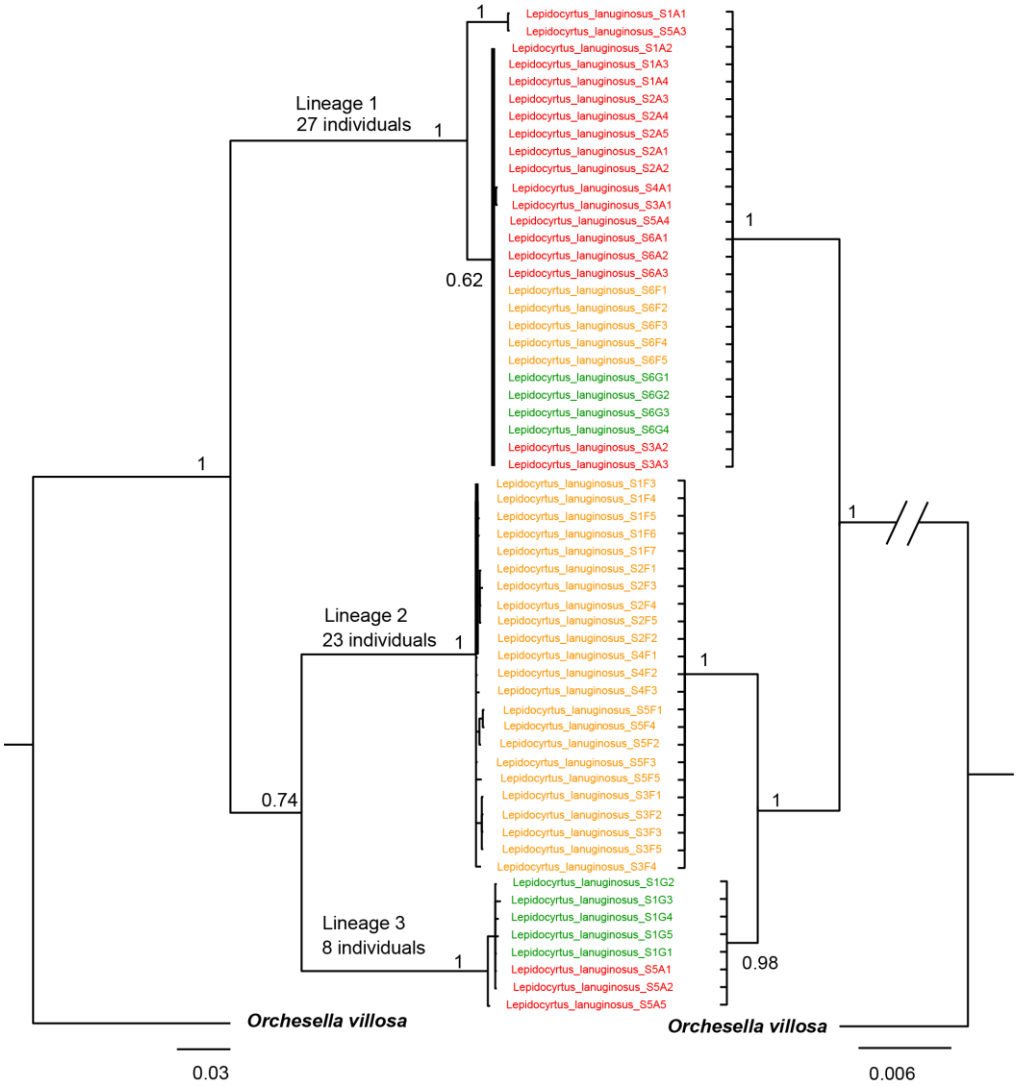


Figure 2.1 Bayesian phylogenetic tree of *Lepidocyrtus lanuginosus* based on single molecular markers: COII (left) and 28S D1-2 (right). Labels of OTUs represent sampling locations (refer to Table 2.1) and habitat types: A, arable land (red); G, grassland (green); F, forest (yellow).

considerably larger than within lineages (Table 2.2). For the mitochondrial markers, mean K2P and uncorrected p-distances within *L. lanuginosus* L1, L2 and L3 were below 1.2%; both nuclear markers were identical within lineages. By contrast, mean K2P between the three *L. lanuginosus* lineages were similar and ranged between 22.2% and 27.5% for both mitochondrial markers (19% and 22.8% for uncorrected p-distances), between 1.3% and 2.8% for 28S D1–2 (1.2% and 2.9% for uncorrected p-distances), and between 1.5% and 5.1% for EF1-α (1.2% and 4.9% for uncorrected p-distances).

Notably, mean K2P distances of mitochondrial markers between *L. cyaneus* and *L. lanuginosus* L1 were 16.8%–17.7%, i.e. about one third lower than between *L. cyaneus* and the other two lineages (22.8%–25.5% for L2 and 21.6%–25.9% for L3). Similarly, genetic distances of the nuclear markers were lowest between *L. cyaneus* and *L. lanuginosus* L1 (0.1% in 28S and 1.5% in EF1-α), but were larger

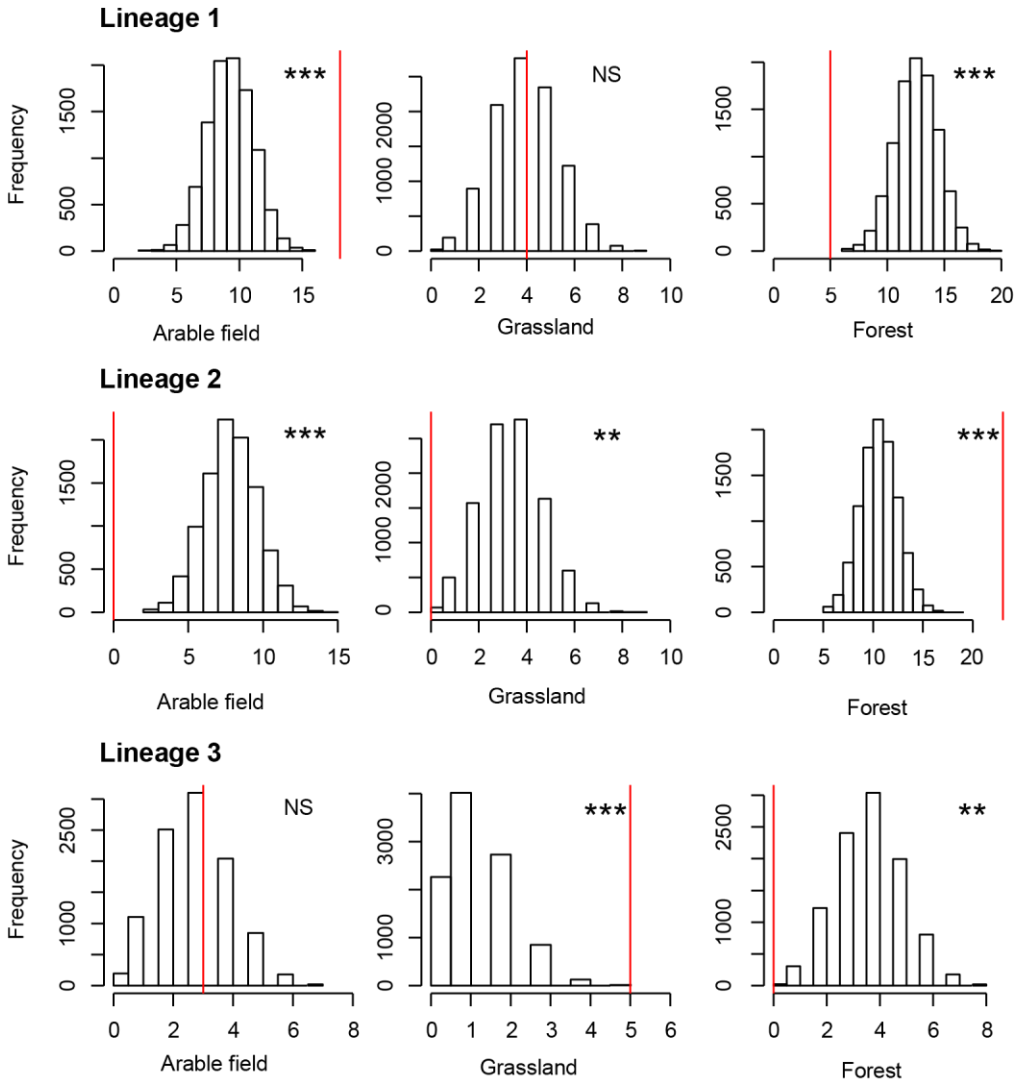


Figure 2.2 Frequency of randomized numbers of individuals of three lineages of *Lepidocyrtus lanuginosus* in the three habitats studied. Red lines indicate observed numbers of individuals; bars show the frequencies of simulated habitat distribution derived from 9,999 permutations. Stars indicate p-values based on the rank of observed number compared to the simulated numbers. ***, $p < 0.001$; **, $p < 0.01$; NS, not significant.

between *L. cyaneus* and *L. lanuginosus* L2 (2.8% in 28S and 4.5% in EF1-α) and L3 (2.5% in 28S D1–2 and 4.6% in EF1-α). Patterns for uncorrected p-distances were similar (Table 2.2). K2P distances of COI between the three lineages of *L. lanuginosus* were larger than 20% (larger than 17.8% for uncorrected p-distances). This resembled distances between the three lineages of *L. lanuginosus* and four other well-defined species of *Lepidocyrtus*, i.e. *L.* cf. *arrabonicus*, *L. paradoxus*, *L.* cf. *violaceus* and *L.* cf. *weidneri* (Table 2.2). By contrast, for COII, 28S D1–2 and EF1-α mean genetic distances between the three lineages of *L. lanuginosus* were lower than the distances between these three lineages and the other four *Lepidocyrtus* species.

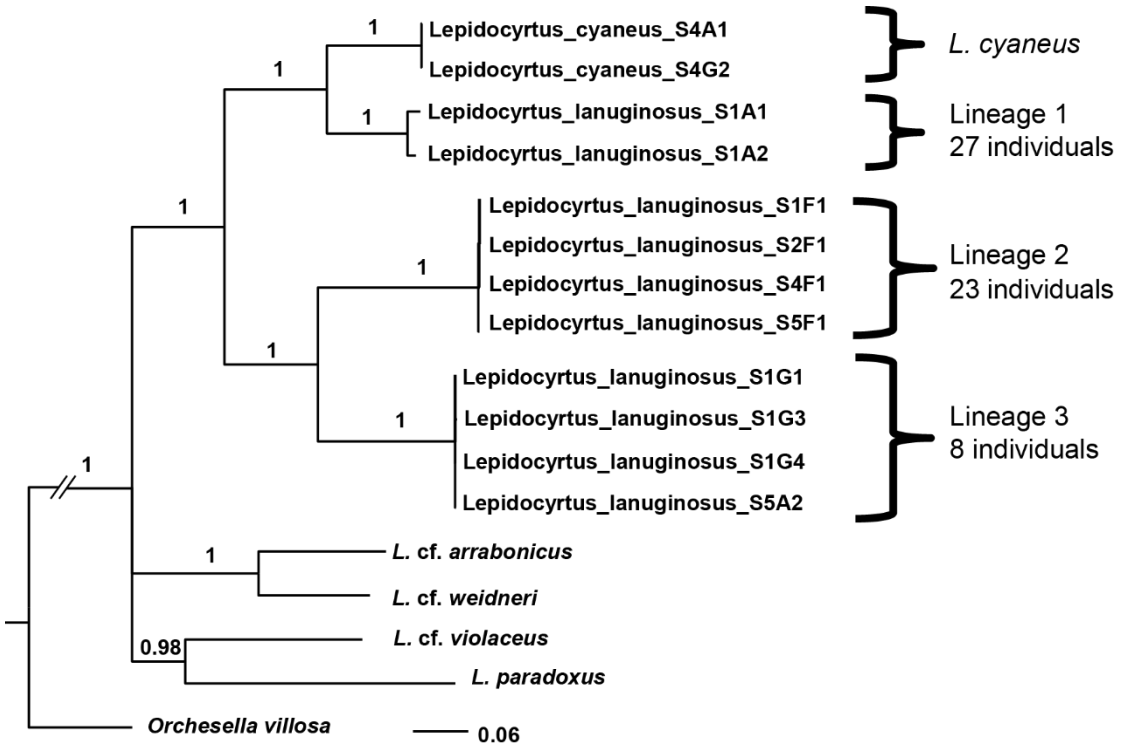


Figure 2.3 Bayesian phylogenetic tree of three lineages of *Lepidocyrtus lanuginosus* and five other species of *Lepidocyrtus* based on concatenated data of four molecular markers: 28S D1-2, elongation factor 1-α, COI and COII. Bayesian posterior probabilities are indicated on nodes. Labels of OTUs represent sampling locations (refer to Table 1) and habitats (A, arable land; G, grassland; F, forest).

The phylogenetic tree based on the combined matrix of the four genes (28S D1–2, EF1-α, COI and COII) show that all lineages of *L. lanuginosus* are monophyletic (Bayesian posterior probabilities (PP) = 1; Figure 2.3). Applying the standard mutation rate of COI for arthropods of 1.5–2.3% sequence divergence per million years (Avise, 1994; Brower, 1994) to the mean K2P distances suggests that the three *L. lanuginosus* lineages diverged 15.9–9.7 million years ago (Mya) in the Miocene (13.4–8.2 Mya based on p-distances).

2.4 Discussion

2.4.1 Habitat sorting

Results of the present study suggest that the morphotype of *L. lanuginosus* does not represent a single monophyletic species but rather they suggest the existence of three distinct genetic lineages in a narrow region in Central Europe. One lineage occurred in each of the three habitat types (L1) and might be regarded as a habitat generalist but with a preference for arable fields. The other two

Table 2.2 Mean inter- and intraspecific K2P and uncorrected p-distances (%) between the three lineages of *Lepidocyrtus lanuginosus* (L1–L3) and between these lineages and five other species of the genus *Lepidocyrtus*. Upper rows represent distances in COI (left panel) and 28S D1–D2 (right panel), lower rows highlighted in grey represent distances in COII (left panel) and elongation factor 1-α (right panel). For each lineage/species the number of individuals analyzed is given (N).

		COI / COII								28S D1-2 / elongation factor 1a							
		1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
K2P distance																	
1	*L. lanuginosus L1* (N=6)	1.1								0							
		1.2								0							
2	*L. lanuginosus L2* (N=6)	23.8	0.4							3.1	0						
		27	0.3							5	0						
3	*L. lanuginosus L3* (N=6)	23.8	22.2	0.2						2.5	1.3	0					
		27.5	23	0.3						5	1.4	0					
4	*L. cyaneus* (N=6)	16.8	22.8	21.6	0					0.1	2.9	2.6	0				
		17.7	25.5	25.9	0					1.5	4.4	4.4	0				
5	*L.* cf. *arrabonicus* (N=6)	23.2	23.6	23.1	22.2	5.8				4.1	4.6	4.4	4.3	0			
		26.2	30	30	27.9	7				9.3	6.5	6	9.3	0			
6	*L. paradoxus (N=5)*	25.8	25.5	26.5	26	24.7	2.8			3.2	3.8	3.2	3.4	3.4	0		
		33.3	37.2	31.9	31	31	1.8			8	8.1	8.6	9.7	10.8	0		
7	*L.* cf. *violaceus* (N=5)	21.7	22.7	23.3	20.8	23.5	22.9	0.2		2.9	4.1	3.5	3.1	3.2	2.9	0	
		30.5	33.9	33.3	30.8	30.4	31.7	0.2		8.4	5.1	5.8	9	5.3	7.4	0	
8	*L.* cf. *weidneri* (N=6)	20.3	22.9	22.1	19.8	20	24.2	21.6	2.1	4.1	4.6	4.4	4.3	0	3.4	3.2	0
		29.6	29.9	29.4	27.4	22.2	32.7	30.6	1.7	9.5	6.8	6.2	9.5	1.2	12.2	6.1	0
p-distance																	
1	*L. lanuginosus L1* (N=6)	1.1								0							
		1.2								0							
2	*L. lanuginosus L2* (N=6)	20.2	0.4							3	0						

		COI / COII								28S D1-2 / elongation factor 1a							
		1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
		22.3	0.3							4.8	0						
3	*L. lanuginosus L3* (N=6)	20.1	19	0.2						2.4	1.3	0					
		22.8	19.6	0.3						4.8	1.4	0					
4	*L. cyaneus* (N=6)	14.8	19.6	18.6	0					0.1	2.9	2.6	0				
		15.4	21.5	21.8	0					1.4	4.3	4.3	0				
5	*L.* cf. *arrabonicus* (N=6)	19.7	20.1	19.6	19	5.5				4	4.4	4.3	4.2	0			
		22	24.4	24.6	23.2	6.5				8.6	6.2	5.7	8.6	0			
6	*L. paradoxus (N=5)*	21.7	21.4	22.1	21.8	20.9	2.7			3.2	3.7	3.2	3.3	3.3	0		
		26.7	29	25.8	25.2	25.2	1.8			7.6	7.6	8.1	9.1	10	0		
7	*L.* cf. *violaceus* (N=5)	18.7	19.4	19.8	18	19.9	19.5	0.2		2.9	4	3.4	3	3.1	2.9	0	
		24.8	26.9	26.6	25	24.7	25.5	0.2		8	4.9	5.6	8.4	5.1	7	0	
8	*L.* cf. *weidneri* (N=6)	17.6	19.5	19	17.2	17.3	20.5	18.6	2	4	4.4	4.3	4.2	0	3.3	3.1	0
		24.3	24.3	24.1	22.8	19	26.3	24.9	1.7	8.9	6.5	6	8.9	1.2	11.2	5.8	0

lineages were restricted either to forests (L2) or to grasslands (L3), suggesting more specialized ecotypes of *L. lanuginosus*. The proximity of the different habitats at the six sites support differential colonization of forest and open habitats (arable fields and grassland) indicating that the different lineages were sorted by environmental factors. The high genetic distances between lineages suggest impeded gene flow either by isolation by distance (van der Wurff et al., 2005) or isolation by environment (Wang and Summers, 2010). Dispersal limitation between habitats is unlikely because *L. lanuginosus* is a mobile and surface active species (Ponge et al., 2006; Salmon et al., 2014), exhibiting high mobility in grasslands as well as forests (Auclerc et al., 2009). Rather, sorting of lineages into different habitats due to lineage-specific preferences or physiological constraints to habitat characteristics is more likely. Forests are stable habitats usually with thick organic layers and high soil carbon content providing a variety of resources (Hopkin, 1997). In contrast, open habitats such as arable fields and grasslands typically are characterized by disturbances due to agricultural practices and stronger fluctuations in temperature and soil moisture (Batlle-Aguilar et al., 2011). Differences in tolerance to disturbances among lineages are possible, but ecophysiological experiments are needed to test this hypothesis. Our results suggest habitat preferences of genetic lineages in *L. lanuginosus* but this needs to be confirmed by more extensive sampling of different habitats over a larger geographic area in particular for the rare genotype *L. lanuginosus* L3.

2.4.2 Genetic distances

High genetic distance between the three *L. lanuginosus* lineages suggests the existence of cryptic species in the collected morphologically coherent specimens. Each of the four genetic markers allowed to separate the three lineages of *L. lanuginosus* and thus can be used as barcoding marker. Inter-lineage K2P distances in COI were larger than 0.21%, which is similar to the inter-lineage distances of cryptic species described by Porco et al., (2012a), and almost twice as high as the suggested boundary of intraspecific variability of Collembola (Anslan and Tedersoo, 2015; Porco et al., 2014). The inter-lineage genetic distances of the nuclear marker 28S D1–2 also exceeded the previously suggested minimum interspecific genetic distance between different Collembola species (Anslan and Tedersoo, 2015). In addition, K2P and uncorrected p-distances in both of the mitochondrial and nuclear markers between the three lineages of *L. lanuginosus* exceeded those between *L. cyaneus* and *L. lanuginosus* L1 emphasizing the existence of cryptic species within *L. lanuginosus*.

Lepidocyrtus lanuginosus is widespread and frequently reported in soil ecological studies, as we obtained 355 hits for this species in Google Scholar for publications before 2016. Cryptic species within

L. lanuginosus without clearly delimiting morphological characters suggest that the diversity of Collembola is underestimated. Our study adds to recent evidence that cryptic species commonly exist within morphospecies of Collembola such as *Heteromurus major* (Moniez, 1889), *Deutonura monticola* (Cassagnau, 1954) (Porco et al., 2012a), and *Parisotoma notabilis* (Schäffer, 1896) (Porco et al., 2012a,b). Generalist Collembola species colonize a variety of habitats (Auclerc et al., 2009; Heiniger et al., 2015). However, except for few studies (Carapelli et al., 1995; Porco et al., 2014) only single habitat types were investigated (Cicconardi et al., 2013; Porco et al., 2012b; Timmermans et al., 2005; von Saltzwedel et al., 2016), thereby disregarding environmental associations of cryptic species. Taking habitat types into account when investigating genetic diversity of ubiquitous Collembola may lead to the discovery of cryptic or new species and provide a more accurate picture of local and regional diversity.

2.4.3 Systematics

According to identification keys (Fjellberg, 2007; Hopkin, 2007; Mateos, 2008) all three lineages in this study unequivocally belong to *L. lanuginosus* with common main characters, differing clearly from *L. cyaneus* in pigmentation. Earlier studies based on COI proposed that color patterns are valid to discriminate *Lepidocyrtus* species in North America (Soto-Adames, 2002). However, as shown in the concatenated phylogenetic tree of this study, the three lineages of *L. lanuginosus* are well separated, at least in Central Europe. At our study sites, *L. cyaneus* is more closely related to *L. lanuginosus* L1 than *L. lanuginosus* L1 to L2 and L3, suggesting that *L. cyaneus* may form part of *L. lanuginosus* complex. Further taxonomic investigations are needed to explore morphological differences of the *L. lanuginosus* lineages found in our study and in the Mediterranean (Cicconardi et al., 2010), e.g. body chaetotaxy (Mateos, 2012). Unequivocal identification of lineages of *L. lanuginosus* with each of the four genetic markers used in this study demonstrates that molecular markers are a reliable and fast tool to detect cryptic species, which may facilitate taxonomic research of Collembola species and the identification of possible species-specific morphological characters.

2.5 Conclusions

Our findings show that even at a narrow geographical scale different genetic lineages or cryptic species of *L. lanuginosus* occur. Lineage sorting by habitat suggests that specimens from different habitats in close vicinity may belong to different cryptic lineages. Genetically the three *L. lanuginosus* lineages are as distant to each other as good species. Future studies need to explore the physiological

characters responsible for habitat sorting of cryptic species / lineages of the *L. lanuginosus* species complex.

Acknowledgements

We thank Jo-Fan Chao for her help in laboratory work. The study was supported by the German Ministry of Education, Science and Technology (BMBF). Bing Zhang was supported by the China Scholarship Council.

References

Anslan, S., Tedersoo, L., 2015. Performance of cytochrome c oxidase subunit I (COI), ribosomal DNA Large Subunit (LSU) and Internal Transcribed Spacer 2 (ITS2) in DNA barcoding of Collembola. Eur. J. Soil Biol. 69, 1–7. doi:10.1016/j.ejsobi.2015.04.001

Auclerc, A., Ponge, J.F., Barot, S., Dubs, F., 2009. Experimental assessment of habitat preference and dispersal ability of soil springtails. Soil Biol. Biochem. 41, 1596–1604. doi:10.1016/j.soilbio.2009.04.017

Avise, J.C., 1994. Molecular Markers, Natural History and Evolution. Chapman & Hall, New York.

Batlle-Aguilar, J., Brovelli, A., Porporato, A., Barry, D. a, 2011. Modelling soil carbon and nitrogen cycles during land use change. A review. Agron. Sustain. Dev. 31, 251–274. doi:10.1051/agro/2010007

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K., Das, I., 2007. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. 22, 148–155. doi:10.1016/j.tree.2006.11.004

Bidochka, M.J., Kamp, A.M., Lavender, T.M., Dekoning, J., De Croos, J.N.A., 2001. Habitat association in two genetic groups of the Insect-pathogenic fungus *Metarhizium anisopliae*: uncovering cryptic species? Appl. Environ. Microbiol. 67, 1335–1342. doi:10.1128/AEM.67.3.1335-1342.2001

Brower, A. V., 1994. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. Proc. Natl. Acad. Sci. 91, 6491–6495. doi:10.1073/pnas.91.14.6491

Carapelli, A., Frati, F., Fanciulli, P.P., Dallai, R., 1995. Genetic differentiation of six sympatric species of *Isotomurus* (Collembola, Isotomidae): is there any difference in their microhabitat preference? Eur. J. Soil Biol. 31, 87–99.

Cicconardi, F., Fanciulli, P.P., Emerson, B.C., 2013. Collembola, the biological species concept and the underestimation of global species richness. Mol. Ecol. 22, 5382–5396. doi:10.1111/mec.12472

Cicconardi, F., Nardi, F., Emerson, B.C., Frati, F., Fanciulli, P.P., 2010. Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. Mol. Ecol. 19, 386–400. doi:10.1111/j.1365-294X.2009.04457.x

D'Haese, C.A., 2002. Were the first springtails semi-aquatic? A phylogenetic approach by means of 28S rDNA and optimization alignment. Proc. R. Soc. B Biol. Sci. 269, 1143–1151. doi:10.1098/rspb.2002.1981

Davidson-Watts, I., Walls, S., Jones, G., 2006. Differential habitat selection by Pipistrellus pipistrellus and Pipistrellus pygmaeus identifies distinct conservation needs for cryptic species of echolocating bats. Biol. Conserv. 133, 118–127. doi:10.1016/j.biocon.2006.05.027

Dufrêne, M., Legendre, P., 1997. Species assemblages and indicator species: The need for a flexible asymmetrical approach. Ecol. Monogr. 67, 345–366. doi:10.2307/2963459

Eisenring, M., Altermatt, F., Westram, A.M., Jokela, J., 2016. Habitat requirements and ecological niche of two cryptic amphipod species at landscape and local scales. Ecosphere 7, 1–13. doi:10.1002/ecs2.1319

Emerson, B.C., Cicconardi, F., Fanciulli, P.P., Shaw, P.J.A., 2011. Phylogeny, phylogeography, phylobetadiversity and the molecular analysis of biological communities. Philos. Trans. R. Soc. B Biol. Sci. 366, 2391–2402. doi:10.1098/rstb.2011.0057

Fjellberg, A., 2007. The Collembola of Fennoscandia and Denmark, Part II: Entomobryomorpha and Symphypleona, Brill. Brill, Boston. doi:10.1163/ej.9789004157705.i-265

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. Mol. Mar. Biol. Biotechnol. 3, 294–299.

Frati, F., Dell'Ampio, E., Casasanta, S., Carapelli, A., Fanciulli, P. Pietro, 2000. Large amounts of genetic divergence among Italian species of the genus Orchesella (Insecta, Collembola) and the relationships of two new species. Mol. Phylogenet. Evol. 17, 456–461. doi:10.1006/mpev.2000.0854

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98. doi:citeulike-article-id:691774

Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B Biol. Sci. 270, 313–321. doi:10.1098/rspb.2002.2218

Heiniger, C., Barot, S., Ponge, J.-F., Salmon, S., Meriguet, J., Carmignac, D., Suillerot, M., Dubs, F., 2015. Collembolan preferences for soil and microclimate in forest and pasture communities. Soil Biol. Biochem. 86, 181–192. doi:10.1016/j.soilbio.2015.04.003

Hogg, I.D., Hebert, P.D.., 2004. Biological identification of springtails (Hexapoda: Collembola) from the Canadian Arctic, using mitochondrial DNA barcodes. Can. J. Zool. 82, 749–754. doi:10.1139/z04-041

Hopkin, S.P., 2007. A Key to the Springtails (Collembola) of Britain and Ireland, Field Studies Council (AIDGAP Project). FSC Publicatio, Shrewsbury.

Hopkin, S.P., 1997. Biology of the springtails: (Insecta: Collembola). Oxford University Press, Oxford, UK.

James, S.W., Porco, D., Decaëns, T., Richard, B., Rougerie, R., Erséus, C., 2010. DNA barcoding reveals cryptic diversity in Lumbricus terrestris L., 1758 (Clitellata): resurrection of L. herculeus (Savigny, 1826). PLoS One 5. doi:10.1371/journal.pone.0015629

Katz, A.D., Giordano, R., Soto-Adames, F.N., 2015. Operational criteria for cryptic species delimitation when evidence is limited, as exemplified by North American *Entomobrya* (Collembola: Entomobryidae). Zool. J. Linn. Soc. 173, 818–840. doi:10.1111/zoj.12220

Kempson, D., Lloyd, M., Ghelardi, R., 1963. A new extractor for woodland litter. Pedobiologia. 3, 1–30.

King, R.A., Tibble, A.L., Symondson, W.O.C., 2008. Opening a can of worms: Unprecedented sympatric cryptic diversity within British lumbricid earthworms. Mol. Ecol. 17, 4684–4698. doi:10.1111/j.1365-294X.2008.03931.x

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701. doi:10.1093/molbev/mss020

Mateos, E., 2012. The European *Lepidocyrtus lanuginosus* group (Collembola: Entomobryidae), definition and description of a new species from Spain. Zootaxa 3570, 69–81.

Mateos, E., 2008. The European *Lepidocyrtus* Bourlet, 1839 (Collembola: Entomobryidae). Zootaxa 1769, 35–59.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, in: 2010 Gateway Computing Environments Workshop (GCE). IEEE, pp. 1–8. doi:10.1109/GCE.2010.5676129

Ponge, J.F., Dubs, F., Gillet, S., Sousa, J.P., Lavelle, P., 2006. Decreased biodiversity in soil springtail communities: the importance of dispersal and landuse history in heterogeneous landscapes. Soil Biol. Biochem. 38, 1158–1161. doi:10.1016/j.soilbio.2005.09.004

Porco, D., Bedos, A., Greenslade, P., Janion, C., Skarzyński, D., Stevens, M.I., Jansen van Vuuren, B., Deharveng, L., 2012a. Challenging species delimitation in Collembola: cryptic diversity among common springtails unveiled by DNA barcoding. Invertebr. Syst. 26, 470–477. doi:10.1071/IS12026

Porco, D., Potapov, M., Bedos, A., Busmachiu, G., Weiner, W.M., Hamra-Kroua, S., Deharveng, L., 2012b. Cryptic diversity in the ubiquist species *Parisotoma notabilis* (Collembola, Isotomidae): a long-used chimeric species? PLoS One 7, e46056. doi:10.1371/journal.pone.0046056

Porco, D., Skarzyński, D., Decaëns, T., Hebert, P.D.N., Deharveng, L., 2014. Barcoding the Collembola of Churchill: A molecular taxonomic reassessment of species diversity in a sub-Arctic area. Mol. Ecol. Resour. 14, 249–261. doi:10.1111/1755-0998.12172

Posada, D., 2008. jModelTest: Phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256. doi:10.1093/molbev/msn083

Querner, P., Bruckner, A., Drapela, T., Moser, D., Zaller, J.G., Frank, T., 2013. Landscape and site effects on Collembola diversity and abundance in winter oilseed rape fields in eastern Austria. Agric. Ecosyst. Environ. 164, 145–154. doi:10.1016/j.agee.2012.09.016

Ramirez-Gonzalez, R., Yu, D.W., Bruce, C., Heavens, D., Caccamo, M., Emerson, B.C., 2013. PyroClean: denoising pyrosequences from protein-coding amplicons for the recovery of interspecific and intraspecific genetic variation. PLoS One 8, 1–11. doi:10.1371/journal.pone.0057615

Ronquist, F., Teslenko, M., van der Mark, 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. Syst. Biol. 61, 539–542. doi:10.1093/sysbio/sys029

Rosenberger, M., Maraun, M., Scheu, S., Schaefer, I., 2013. Pre- and post-glacial diversifications shape genetic complexity of soil-living microarthropod species. Pedobiologia. 56, 79–87. doi:10.1016/j.pedobi.2012.11.003

Salmon, S., Ponge, J.F., Gachet, S., Deharveng, L., Lefebvre, N., Delabrosse, F., 2014. Linking species, traits and habitat characteristics of Collembola at European scale. Soil Biol. Biochem. 75, 73–85. doi:10.1016/j.soilbio.2014.04.002

Schäffer, S., Pfingstl, T., Koblmüller, S., Winkler, K.A., Sturmbauer, C., Krisper, G., 2010. Phylogenetic analysis of European Scutovertex mites (Acari, Oribatida, Scutoverticidae) reveals paraphyly and cryptic diversity: A molecular genetic and morphological approach. Mol. Phylogenet. Evol. 55, 677–688. doi:10.1016/j.ympev.2009.11.025

Soto-Adames, F.N., 2002. Molecular phylogeny of the Puerto Rican *Lepidocyrtus* and *Pseudosinella* (Hexapoda: Collembola), a validation of Yoshii's "color pattern species." Mol. Phylogenet. Evol. 25, 27–42. doi:10.1016/S1055-7903(02)00250-6

Stevens, M.I., Frati, F., McGaughran, A., Spinsanti, G., Hogg, I.D., 2007. Phylogeographic structure suggests multiple glacial refugia in northern Victoria Land for the endemic Antarctic springtail *Desoria klovstadi* (Collembola, Isotomidae). Zool. Scr. 36, 201–212. doi:10.1111/j.1463-6409.2006.00271.x

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. Mol. Biol. Evol. 28, 2731–2739. doi:10.1093/molbev/msr121

Tarjuelo, I., Posada, D., Crandall, K.A., Pascual, M., Turon, X., 2001. Cryptic species of Clavelina (Ascidiacea) in two different habitats: Harbours and rocky littoral zones in the northwestern Mediterranean. Mar. Biol. 139, 455–462. doi:10.1007/s002270100587

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680. doi:10.1093/nar/22.22.4673

Timmermans, M.J.T.N., Ellers, J., Mariën, J., Verhoef, S.C., Ferwerda, E.B., van Straalen, N.M., 2005. Genetic structure in Orchesella cincta (Collembola): Strong subdivision of European populations inferred from mtDNA and AFLP markers. Mol. Ecol. 14, 2017–2024. doi:10.1111/j.1365-294X.2005.02548.x

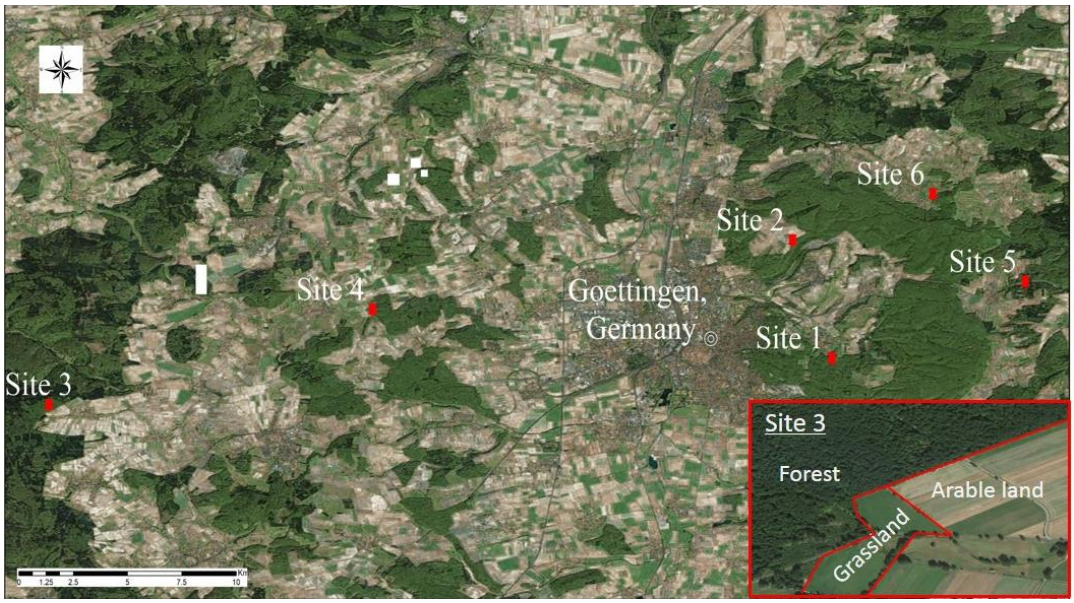
van der Wurff, A.W.G., Gols, R., Ernsting, G., van Straalen, N.M., 2005. Population genetic structure of *Orchesella cincta* (Collembola; Hexapoda) in NW Europe, as revealed by microsatellite markers. Pedobiologia. 49, 167–174. doi:10.1016/j.pedobi.2004.10.004

von Saltzwedel, H., Scheu, S., Schaefer, I., 2017. Genetic structure and distribution of *Parisotoma notabilis* (Collembola) in Europe: Cryptic diversity, split of lineages and colonization patterns. PLoS One 12, e0170909. doi:10.1371/journal.pone.0170909

von Saltzwedel, H., Scheu, S., Schaefer, I., 2016. Founder events and pre-glacial divergences shape the genetic structure of European Collembola species. BMC Evol. Biol. 16, 148. doi:10.1186/s12862-016-0719-8

Wang, I.J., Summers, K., 2010. Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. Mol. Ecol. 19, 447–458. doi:10.1111/j.1365-294X.2009.04465.x

Supplementary Materials



Supplementary Figure S2.1 Sampling sites of *Lepidocyrtus cyaneus* and *L. lanuginosus* in Central Europe near Göttingen, Germany. The map was generated with ArcGIS version 10.0. Inlet: Location of the three habitat types shown at Site 3 as example.

Supplementary Table S2.1 Collembola specimens used for DNA barcoding with accession numbers of sequences of 28S D1–2, COII, COI and elongation factor 1-α (EF1-α) in GenBank.

Species	Individual code	Site No.	Village name	Habitat	28S D1–D2	COII	COI	EF1-α
Lepidocyrtus cyaneus	Lepidocyrtus_cyaneus_S4A1	4	Ossenfeld	arable land	MH160158	MH144023	MH143938	MH178011
	Lepidocyrtus_cyaneus_S4G1	4	Ossenfeld	grassland	MH160159	MH144024	MH143939	MH178012
	Lepidocyrtus_cyaneus_S4G2	4	Ossenfeld	grassland	MH160160	MH144025	MH143940	MH178013
	Lepidocyrtus_cyaneus_S5A1	5	Waake	arable land	MH160161	MH144026	MH143941	MH178014
	Lepidocyrtus_cyaneus_S5A2	5	Waake	arable land	MH160162	MH144027	MH143942	MH178015
	Lepidocyrtus_cyaneus_S5A3	5	Waake	arable land	MH160163	MH144028	MH143943	MH178016
Lepidocyrtus lanuginosus	Lepidocyrtus_lanuginosus_S1A1	1	Herberhausen	arable land	MH160100	MH143965	MH143920	MH177993
	Lepidocyrtus_lanuginosus_S1A2	1	Herberhausen	arable land	MH160101	MH143966	MH143921	MH177994
	Lepidocyrtus_lanuginosus_S1A3	1	Herberhausen	arable land	MH160102	MH143967		
	Lepidocyrtus_lanuginosus_S1A4	1	Herberhausen	arable land	MH160103	MH143968		
	Lepidocyrtus_lanuginosus_S1F3	1	Herberhausen	forest	MH160104	MH143969		
	Lepidocyrtus_lanuginosus_S1F4	1	Herberhausen	forest	MH160105	MH143970		
	Lepidocyrtus_lanuginosus_S1F5	1	Herberhausen	forest	MH160106	MH143971		
	Lepidocyrtus_lanuginosus_S1F6	1	Herberhausen	forest	MH160107	MH143972	MH143922	MH178004
	Lepidocyrtus_lanuginosus_S1F7	1	Herberhausen	forest	MH160108	MH143973		
	Lepidocyrtus_lanuginosus_S1G2	1	Herberhausen	grassland	MH160109	MH143974	MH143924	MH177996
	Lepidocyrtus_lanuginosus_S1G3	1	Herberhausen	grassland	MH160110	MH143975	MH143925	MH177997
	Lepidocyrtus_lanuginosus_S1G4	1	Herberhausen	grassland	MH160111	MH143976	MH143926	MH177998
	Lepidocyrtus_lanuginosus_S1G5	1	Herberhausen	grassland	MH160112	MH143977		
	Lepidocyrtus_lanuginosus_S1G6	1	Herberhausen	grassland	MH160113	MH143978	MH143923	MH177995
	Lepidocyrtus_lanuginosus_S2A3	2	Deppoldshausen	arable land	MH160114	MH143979	MH143927	MH177999
	Lepidocyrtus_lanuginosus_S2A4	2	Deppoldshausen	arable land	MH160115	MH143980		
	Lepidocyrtus_lanuginosus_S2A5	2	Deppoldshausen	arable land	MH160116	MH143981		
	Lepidocyrtus_lanuginosus_S2A6	2	Deppoldshausen	arable land	MH160117	MH143982		
	Lepidocyrtus_lanuginosus_S2A7	2	Deppoldshausen	arable land	MH160118	MH143983		

Species	Individual code	Site No.	Village name	Habitat	28S D1–D2	COII	COI	EF1-α
	Lepidocyrtus_lanuginosus_S2F1	2	Deppoldshausen	forest	MH160119	MH143984		
	Lepidocyrtus_lanuginosus_S2F2	2	Deppoldshausen	forest	MH160120	MH143985		
	Lepidocyrtus_lanuginosus_S2F3	2	Deppoldshausen	forest	MH160121	MH143986		
	Lepidocyrtus_lanuginosus_S2F4	2	Deppoldshausen	forest	MH160122	MH143987	MH143928	MH178002
	Lepidocyrtus_lanuginosus_S2F5	2	Deppoldshausen	forest	MH160123	MH143988	MH143929	MH178003
	Lepidocyrtus_lanuginosus_S3A1	3	Ellershausen	arable land	MH160124	MH143989		
	Lepidocyrtus_lanuginosus_S3A2	3	Ellershausen	arable land	MH160125	MH143990		
	Lepidocyrtus_lanuginosus_S3A3	3	Ellershausen	arable land	MH160126	MH143991		
	Lepidocyrtus_lanuginosus_S3F1	3	Ellershausen	forest	MH160127	MH143992		
	Lepidocyrtus_lanuginosus_S3F2	3	Ellershausen	forest	MH160128	MH143993		
	Lepidocyrtus_lanuginosus_S3F3	3	Ellershausen	forest	MH160129	MH143994		
	Lepidocyrtus_lanuginosus_S3F4	3	Ellershausen	forest	MH160130	MH143995		
	Lepidocyrtus_lanuginosus_S3F5	3	Ellershausen	forest	MH160131	MH143996		
	Lepidocyrtus_lanuginosus_S4A1	4	Ossenfeld	arable land	MH160132	MH143997		
	Lepidocyrtus_lanuginosus_S4F1	4	Ossenfeld	forest	MH160133	MH143998	MH143930	MH178005
	Lepidocyrtus_lanuginosus_S4F2	4	Ossenfeld	forest	MH160134	MH143999		
	Lepidocyrtus_lanuginosus_S4F3	4	Ossenfeld	forest	MH160135	MH144000		
	Lepidocyrtus_lanuginosus_S5A1	5	Waake	arable land	MH160136	MH144001	MH143931	MH178000
	Lepidocyrtus_lanuginosus_S5A2	5	Waake	arable land	MH160137	MH144002		
	Lepidocyrtus_lanuginosus_S5A3	5	Waake	arable land	MH160138	MH144003		
	Lepidocyrtus_lanuginosus_S5A4	5	Waake	arable land	MH160139	MH144004		
	Lepidocyrtus_lanuginosus_S5A5	5	Waake	arable land	MH160140	MH144005	MH143932	MH178001
	Lepidocyrtus_lanuginosus_S5F1	5	Waake	forest	MH160141	MH144006	MH143933	MH178006
	Lepidocyrtus_lanuginosus_S5F2	5	Waake	forest	MH160142	MH144007		
	Lepidocyrtus_lanuginosus_S5F3	5	Waake	forest	MH160143	MH144008		
	Lepidocyrtus_lanuginosus_S5F4	5	Waake	forest	MH160144	MH144009	MH143934	MH178007
	Lepidocyrtus_lanuginosus_S5F5	5	Waake	forest	MH160145	MH144010		
	Lepidocyrtus_lanuginosus_S6A1	6	Billingshausen	arable land	MH160146	MH144011		

Species	Individual code	Site No.	Village name	Habitat	28S D1–D2	COII	COI	EF1-α
	Lepidocyrtus_lanuginosus_S6A2	6	Billingshausen	arable land	MH160147	MH144012		
	Lepidocyrtus_lanuginosus_S6A3	6	Billingshausen	arable land	MH160148	MH144013	MH143935	MH178008
	Lepidocyrtus_lanuginosus_S6F1	6	Billingshausen	forest	MH160149	MH144014	MH143936	MH178009
	Lepidocyrtus_lanuginosus_S6F2	6	Billingshausen	forest	MH160150	MH144015		
	Lepidocyrtus_lanuginosus_S6F3	6	Billingshausen	forest	MH160151	MH144016		
	Lepidocyrtus_lanuginosus_S6F4	6	Billingshausen	forest	MH160152	MH144017		
	Lepidocyrtus_lanuginosus_S6F5	6	Billingshausen	forest	MH160153	MH144018		
	Lepidocyrtus_lanuginosus_S6G1	6	Billingshausen	grassland	MH160154	MH144019	MH143937	MH178010
	Lepidocyrtus_lanuginosus_S6G2	6	Billingshausen	grassland	MH160155	MH144020		
	Lepidocyrtus_lanuginosus_S6G3	6	Billingshausen	grassland	MH160156	MH144021		
	Lepidocyrtus_lanuginosus_S6G4	6	Billingshausen	grassland	MH160157	MH144022		
Lepidocyrtus cf. *arrabonicus*	LEarrabonicus_Leb_1	7	Lebeny	grassland	MH160164	MH144029	MH143945	MH178017
	LEarrabonicus_Leb_2	8	Lebeny	grassland	MH160165	MH144030	MH143946	MH178018
	LEarrabonicus_Lebe_2	8	Lebeny	grassland	MH160166	MH144031	MH143944	MH178019
	LEarrabonicus_Mpa_1	9	Mantova Park	grassland	MH160167	MH144032	MH143947	MH178020
	LEarrabonicus_Niea_2	10	Nitra East	grassland	MH160168	MH144033	MH143948	MH178021
	LEarrabonicus_Niea_3	10	Nitra East	grassland	MH160169	MH144034	MH143949	MH178022
Lepidocyrtus paradoxus	LEparadoxus_Goe_1	2	Goettingen	grassland	MH160170	MH144035	MH143950	
	LEparadoxus_Goe_5	2	Goettingen	grassland	MH160171	MH144036	MH143951	
	LEparadoxus_Mas_1	11	Masetti	grassland	MH160172	MH144037	MH143952	MH178032
	LEparadoxus_Ott_1	12	Otterbach	grassland	MH160173	MH144038	MH143953	
	LEparadoxus_Sba_1	12	Sala Baganza	grassland	MH160174	MH144039	MH143954	MH178033
Lepidocyrtus cf. *violaceus*	LEviolaceus_Pga_1	13	Park Gardasee	grassland	MH160175	MH144040	MH143955	MH178023
	LEviolaceus_Pga_2	13	Park Gardasee	grassland	MH160176	MH144041	MH143956	
	LEviolaceus_Pga_3	13	Park Gardasee	grassland	MH160177	MH144042	MH143957	MH178024
	LEviolaceus_Pga_4	13	Park Gardasee	grassland	MH160178	MH144043	MH143958	MH178025
	LEviolaceus_Pga_5	13	Park Gardasee	grassland	MH160179	MH144044		
	LEweidneri_Brn_1	14	Brno	grassland	MH160180	MH144045	MH143959	MH178026

Species	Individual code	Site No.	Village name	Habitat	28S D1–D2	COII	COI	EF1-α
Lepidocyrtus cf. *weidneri*	LEweidneri_Brn_2	14	Brno	grassland	MH160181	MH144046	MH143960	MH178027
	LEweidneri_Brn_4	14	Brno	grassland	MH160182	MH144047	MH143961	MH178028
	LEweidneri_Goe_1	1	Goettingen	grassland	MH160183	MH144048	MH143962	MH178029
	LEweidneri_Put_2	15	Putlitz	grassland	MH160184	MH144049	MH143963	MH178030
	LEweidneri_Put_4	15	Putlitz	grassland	MH160185	MH144050	MH143964	MH178031

Chapter III

DNA-based approaches uncover cryptic diversity in the European *Lepidocyrtus lanuginosus* species group (Collembola: Entomobryidae)

Bing Zhang, Ting-Wen Chen, Eduardo Mateos, Stefan Scheu, Ina Schaefer

Abstract

DNA sequence data and phylogenies are useful tools for species delimitation, especially in taxa comprising cryptic species. The *Lepidocyrtus lanuginosus* species group (Collembola: Entomobryidae) comprises three morphospecies and distinct cryptic species. We applied three DNA-based methods to delimit species boundaries of the three morphospecies of the *L. lanuginosus* species group across Central and Southern Europe. Based on COI and COII we identified gaps of genetic distances that indicate species boundaries, and found ten and nine distinct genetic lineages in *L. cyaneus* and *L. lanuginosus*, respectively. The nuclear gene elongation factor 1-α (EF1-α) delimited 89% of the lineages but 28S rDNA (D1–2 domain) was too conserved for this purpose. The phylogenetic trees showed that *L. cyaneus* and *L. lanuginosus* are polyphyletic, suggesting that body color is insufficient for delimiting species in the *L. lanuginosus* species group. Our study challenges the current morphology-based species delimitation in the *L. lanuginosus* species group and suggests that molecular approaches are needed for fast and accurate determination of Collembola species in both taxonomic and ecological studies. Overall, the results suggest that wide geographic sampling combined with molecular phylogenetic approaches is needed to delimit species and to understand the full range of cryptic diversity in Collembola.

Keywords: springtail; color; barcoding; ribosomal subunit 28S rDNA D1–2 domain; elongation factor 1-α; cytochrome c oxidase subunit I and II

3.1 Introduction

Accurate species identification and assessment of species diversity is essential for a wide range of biological studies because species differ in their evolutionary history, adaptation and ecological function (Carstens et al., 2013; Struck et al., 2017). Various operational criteria are used for species identification depending on the species concept invoked e.g., the biological, ecological, evolutionary, or phylogenetic species concept (de Queiroz, 2007, Wiens, 2007). Incompatible species concepts and their associated definitions can lead to different conclusions concerning the boundaries and number of species. Morphological data have been largely used to delimit and describe species, but morphologically coherent species may comprise cryptic species (Bickford et al., 2007). Analyses of DNA sequences demonstrated that cryptic species exist in a range of taxa, i.e., genetically distinct lineages exist in morphologically undifferentiated species resulting in an unregistered biodiversity (Feulner et al., 2006; Grundt et al., 2006; Hebert et al., 2004).

Due to considerable morphological conservatism, species delimitation is particularly difficult in Collembola, which are one of the most numerous soil animals on earth and occur in virtually all terrestrial ecosystems and habitats (Hopkin, 1997). Some morphology-based Collembola species comprise high levels of cryptic diversity as indicated by molecular studies (Cicconardi et al., 2013, 2010; Porco et al., 2012b; von Saltzwedel et al., 2017). Among Collembola, the genus *Lepidocyrtus* Bourlet, 1839 comprises approximately 300 species (Bellinger *et al.* 1996-2018), which in Europe have been ascribed to five monophyletic groups based on chaetotaxy and molecular phylogenetic analysis (Mateos et al., 2018). However, DNA-based analyses indicated the existence of cryptic species in the *L. lanuginosus* species group, showing that genetically distinct lineages of *L. lanuginosus* (Gmelin, 1788) occur in arable land, forest and grassland (B. Zhang et al., 2018), and exist in different forest types and regions of the Mediterranean basin (Cicconardi et al., 2010).

Coloration has been used as character for delimiting species in Collembola since the 19th century (Rusek, 2002), and the use of color as valid character is well supported in the family Entomobryidae (Ding et al., 2018; Soto-Adames, 2002), but similar body color pattern exist among distinct species (Katz et al., 2015b). The *L. lanuginosus* species group (Mateos, 2012) comprises three morphospecies which are separated by body color patterns: (1) the body of *L. lanuginosus* is yellow and lacks dark blue pigmentation, (2) the body of *L. bicoloris* Mateos, 2012 is partially dark blue pigmented, and (3) the body of *L. cyaneus* Tullberg, 1871 is entirely dark blue pigmented. Even though additional morphological characters (i.e., seta Fe4 and E4p of the fourth abdominal segment) differ between *L. cyaneus* and *L. lanuginosus* (Mateos, 2012), color still is the main character to differentiate species of the *L. lanuginosus* species group in determination keys (Fjellberg, 2007; Hopkin, 2007; Mateos, 2008a). Using these keys *L. cyaneus* and *L. lanuginosus* are reported as being abundant and widely distributed springtails in Europe (Salmon et al., 2014). However, high intra- and interspecific genetic variation and phylogenetic reconstruction of species indicated the non-monophyly of the morphospecies *L. lanuginosus* (Mateos et al., 2018; B. Zhang et al., 2018), challenging the validity of body color as character to identify species of the *L. lanuginosus* species group.

Methods that use gaps of genetic distances and gene trees to infer species boundaries have been developed to improve the evaluation of true and cryptic species diversity (Knowles and Carstens, 2007; Wiens, 2007). Automatic Barcode Gap Discovery (ABGD) is a fast and simple method which splits sequence alignment datasets into candidate species based on a barcode gap (Puillandre et al., 2012). This method has been used for identification of species in Collembola, e.g., *Protaphorura* and *Coecobrya*. However, a universal threshold of genetic distance for species delimitation in these genera was not recommended, but rather that barcoding gaps or species boundaries should be based on calculations within specific groups (Sun et al., 2017; F. Zhang et al., 2018). Methods using single or

multiple markers and consider evolutionary models improved molecular species delimitation (Hailer et al., 2012; Knowles and Carstens, 2007). Recently developed methods for species delimitation based on multispecies coalescent (MSC) models, such as the Poisson Tree Processes model (PTP, Zhang *et al.* 2013) and Bayesian delimitation of species (Bayesian Phylogenetics and Phylogeography, BPP; Yang and Rannala 2010), also improved molecular delimitation of species, including Collembola, using either single or multi-locus sequence data (Yang and Rannala, 2017; F. Zhang et al., 2018).

In this study we delimited molecular species boundaries in the *L. lanuginosus* species group by analyzing genetic distances among 38 populations of *L. cyaneus* and *L. lanuginosus* across Central and Southern Europe. The sampling included *L. bicoloris* from the locality in Spain where its holotype was described (Mateos, 2012) and three lineages of *L. lanuginosus* from central Germany (B. Zhang et al., 2018). We checked if species diversity delimited by morphology is consistent with that delimited by genes. The three above mentioned DNA-based methods ABGD, PTP, and BPP were applied to delimit genetic boundaries using mitochondrial COI and COII. COI is widely used as barcoding marker for Collembola (and soil animals in general, Rougerie *et al.* 2009), while COII often has been used to study intraspecific genetic variation in European *Lepidocyrtus* species (Cicconardi et al., 2013, 2010; Mateos et al., 2018; B. Zhang et al., 2018). Nuclear 28S D1–2 and EF1-α were also proposed as potential species markers in Collembola (Anslan and Tedersoo, 2015; B. Zhang et al., 2018; Zhang et al., 2014), thus we also evaluated their efficiency as species marker in the *L. lanuginosus* group by analyzing inter- and intra-specific genetic divergence. Another aim of our study was to analyze the validity of body color as morphological character to distinguish the three species by investigating phylogenetic relationships within and between morphospecies. For that purpose we reconstructed phylogenetic trees based on the four above mentioned genetic markers using Bayesian Inference and Maximum Likelihood methods. We hypothesized that (1) the three morphospecies in the *L. lanuginosus* group are genetically well-separated, with the genetic divergence within morphospecies being lower than between morphospecies, and (2) body color is a valid species character reflecting monophyly of species of the *L. lanuginosus* species group.

3.2 Material and methods

3.2.1 Sampling and DNA extraction

Specimens were collected from 34 localities in Southern and Central Europe (supplementary Table S3.1). Specimens in grasslands and arable fields were collected using a modified leaf blower functioning as aspirator (diameter of opening 14 cm), while specimens in forests were collected from litter and soil, and extracted by heat with temperature gradually increasing from 25 to 55°C during 10

days (Kempson et al., 1963). Animals were preserved in 96% ethanol and stored at -20°C until further analyses. Individuals of *Lepidocyrtus* were sorted under a stereo microscope using the key of Hopkin (2007). A total of 91 individuals of *L. cyaneus*, 89 individuals of *L. lanuginosus*, and five individuals of *L. bicoloris* were analyzed. Additionally, one individual of *L. paradoxus* Uzel, 1891 and *L.* cf*. violaceus* Lubbock, 1873 were sorted and used as outgroups in phylogenetic analyses (supplementary Table S3.1). The head of the specimens was cut off and placed on a microscope slide for microscopic inspection of chaetotaxy and the body was used for DNA extraction. After DNA extraction body cuticles were rescued, rinsed with 96% ethanol and mounted on the microscope slide with the respective head. Then, head and body cuticles were inspected under a phase contrast microscope (Zeiss Axio Scope. A1, Jena, Germany) for cephalic and dorsal chaetotaxy using the key of Mateos (2008). Cuticles and heads are kept as vouchers and deposited in the collections of the J.F. Blumenbach Institute of Zoology and Anthropology, University of Göttingen, Germany.

Genomic DNA was extracted using the DNeasy® Blood and Tissue Kit (Qiagen, Hilden, Germany), following the manufacturer's protocol. Purified DNA was eluted in 30 µl HPLC water for PCR. PCR reactions of COI, COII, EF1-α and 28S rDNA comprised of 12.5 µl SuperHot Taq Mastermix (Genaxxon Bioscience GmbH, Ulm, Germany), 1.5 µl each of the respective forward and reverse primers (10 pM), 3 µl H_2O, 1.5 µl $MgCl_2$ (25 mM) and 5 µl template DNA. Primer sequences, fragment lengths and references are summarized in supplementary Table S3.2. PCR conditions included one initial activation step at 95˚C for 15 min, followed by 35 amplification cycles of denaturation at 94˚C (COI and 28S D1–2) or 95˚C (COII and EF1-α) for 30 s, annealing at 45˚C (COI) or 56˚C (COII and EF1-α) for 30 s or 50˚C (28S D1–2) for 45 s and elongation at 72˚C for 30 s, and a final elongation step at 72˚C for 6 min. PCR products were purified using the Genaxxon PCR Purification Kit (Genaxxon Bioscience GmbH, Ulm, Germany) following the manufacturer's protocol, eluted in 30 µl HPLC water and sent to the Göttingen Genome Laboratory (Institute for Microbiology and Genetics, University of Göttingen) for sequencing. All sequences are available at NCBI GenBank and all accession numbers are provided in supplementary Table S3.1.

3.2.2 DNA-based species delimitation

Sequences were checked by eye aided by the chromatographs and ambiguous positions were corrected in Sequencher v5.0 (Gene Codes Corporation, Ann Arbor, MI, USA). Nucleotide sequences were aligned separately for each genetic marker using ClustalW (Thompson et al., 1994) implemented in BioEdit v7.2.5 (Hall, 1999).

Following Zhang *et al.* (2018b), single-locus discovery approaches were employed to generate initial species hypotheses, followed by a multilocus validation approach, which was used to assess the support for the initial grouping. Both distance- and evolutionary model-based methods were applied for the single mitochondrial marker datasets of COI and COII without *a priori* species hypothesis.

Automatic Barcode Gap Discovery (ABGD) clustered sequences into candidate species by detecting barcoding gaps using a command-line version (Puillandre et al., 2012). Prior intraspecific divergence varied from 0.001 (Pmin, a single nucleotide difference) to 0.15 (Pmax). Relative gap width was set to 1.5, with 10 recursive steps, 40 bids for the graphic histogram of distances, K2P model for distance calculation and remaining parameters were set as default.

The Poisson Tree Processes model (PTP) was used for testing species boundaries on non-ultrametric phylogenies by inspecting difference in the number of substitutions between and within species (Zhang et al., 2013). A maximum likelihood (ML) tree was generated for COI and COII in RAxML under the GTR+Γ model. Both single- and multi-rate PTP models were performed in mptp v0.2.3 (Kapli et al., 2017). Bayesian support values were calculated for each clade with a total of 10 million generations, with a burn-in discarding the first 20%.

Consensus molecular operational taxonomic units (MOTUs) based on the results of the ABGD and PTP analyses were designated as candidate species for the subsequent validation approach, i.e. Bayesian species delimitation (BPP). This analyzes multilocus data under the multispecies coalescent model (Rannala and Yang, 2003) using a reversal jump Markov Chain Monte Carlo (rjMCMC) algorithm (Rannala and Yang, 2013; Yang and Rannala, 2010). It also calculates the posterior probabilities of species assignments given a user-specified guide tree. The input species tree was estimated using BEAST v1.8.4 (Heled and Drummond, 2010) based on all four loci. Gene partitions were unlinked and site models were HKY+Y (28S D1–2), GTR+I+G (COI and COII) and HKY+G (EF1-α). We used a Yule speciation prior and an uncorrelated lognormal relaxed clock and a Markov chain Monte Carlo (MCMC) run of 10^8 generations with sampling after 10^4 generations. Effective sample size (ESS) of parameters sampled from the MCMC chain were checked in Tracer 1.6 (Rambaut et al., 2013). Each specimen was *a priori* assigned to species based on the consensus results of the ABGD and PTP analyses, remaining parameters were set as in previous analyses of the ultrametric tree. Each analysis was repeated twice using different starting seeds in BP&P v3.3 using algorithm 0 and fine-tune ε = 2, and was run for 220,000 generations with the first 20,000 as burnin, and samples taken every 5 generations. We used three different combinations of prior distributions for the ancestral population size (θ) and root age (τ0), as proposed by Leaché and Fujita (2010). These combinations assumed (i) large population sizes θ~G(1, 10) and deep divergences τ0~G(1, 10), (ii) small population sizes θ~G(2, 2000) and shallow

divergences τ0~G(2, 2000), and (iii) large population sizes θ~G(1, 10) and relatively shallow divergences τ0~G(2, 2000).

3.2.3 Data analyses

Genetic Kimura-2 parameter (K2P) distances and uncorrected p-distances of each marker within and between lineages were calculated in Mega 5 with pairwise deletion for gaps (Tamura et al., 2011). The four genetic markers were concatenated using SequenceMartrix v1.7.8 (Vaidya et al., 2011), with a final length of 2,251 bp including COI (651 bp), COII (681 bp), EF1-α (210 bp) and 28S rDNA (709 bp) sequences. *Lepidocyrtus paradoxus* and *L.* cf. *violaceus* were used as outgroup for phylogenetic analyses. Identical sequences were removed in all likelihood methods. All three coding positions of the protein-coding genes were included (COI, COII and EF1-α divided into three partitions – 1st, 2nd, and 3rd codons) in the analyses. Best-fitting substitution models were assessed for each locus (partition) under the BIC criterion in PartitionFinder v2.1.1 (Lanfear *et al.* 2017, supplementary Table S3.3).

The partitioned dataset was analyzed by both ML and Bayesian Inference (BI) methods on online CIPRES services (Miller et al., 2010). Maximum Likelihood trees were reconstructed in RAxML v8.1.2 (Stamatakis, 2014) with the GTRGAMMA model for all subsets and 1,000 bootstrap replicates. Bayesian Inference analyses were done in MrBayes v3.2.6 (Ronquist et al., 2012) with different models for eight subsets (supplementary Table S3.3). Model parameters were unlinked and the model allowed the overall rate across partitions to be different. To avoid branch-length overestimation, the compound Dirichlet priors "brlenspr=unconstrained: gammadir (1, 1, 1, 1)" for branch lengths was incorporated (Zhang et al., 2012). The number of generations for the total analysis was set to 50 million, with the chain being sampled every 5,000 generations. The burn-in value was 25% and remaining parameters were set as default. To confirm convergence, the average standard deviation of split frequencies (ASDSF) and the potential scale reduction factor (PSRF) values were visualized in MrBayes, and effective sample size (ESS) values were checked in Tracer 1.6 (Rambaut et al., 2013).

3.3 Results

3.3.1 MOTUs / cryptic species delimitation based on mitochondrial markers

In total, we obtained 148 COI and 184 COII sequences from 185 individuals of the three morphospecies of the *L. lanuginosus* species group (supplementary Table S3.1). The three morphospecies comprised 19 MOTUs / cryptic species, corresponding to ≈0.07 divergence threshold, which were congruent across ABGD and BPP delimitations (Figure 3.1). Ten MOTUs / cryptic species

(lineages 2–11, L2–11) were assigned to *L. cyaneus* and nine (L1 and L12–19) to *L. lanuginosus*. *Lepidocyrtus bicoloris* was grouped to L1 of the *L. lanuginosus* species group.

ABGD generated 19 and 25 initial partitions / MOTUs and 19–55 and 25–58 recursive partitions for COI and COII, respectively, while prior intraspecific divergences (P) varied from 0.001 to 0.0493 (Figure 3.2b, d) for both COI and COII. Rough "barcoding gaps" of COI were observed at K2P distance of 0.055–0.095 (Figure 3.2a), corresponding to plateaus of 19 MOTUs (Figure 3.2b). Barcoding gaps of COII were observed at K2P distance of 0.02–0.03 and 0.06–0.115 (Figure 3.2c), corresponding to plateaus of 25 and 19 MOTUs (Figure 3.2d). Initial and recursive partitions were identical at 19 MOTUs for both COI and COII.

Both single- and multi-rate PTP based on concatenated dataset of COI and COII estimated 20 MOTUs (Figure 3.1), and divided L10 generated by ABGD into two partitions by isolating individual LEcyaneus_Rap_2 from the others. BPP analyses validated a nineteen-species hypothesis for all three prior distributions of ancestral population size and root age with high speciation probabilities (PP: 0.96–1). In the following sections, we refer to the term "lineage" to represent each MOTU / cryptic species in the *L. lanuginosus* species group.

3.3.2 Genetic distances of both nuclear and mitochondrial markers

The mean K2P and uncorrected p-distances of each of the four genetic markers within morphospecies of *L. cyaneus* (10 lineages) and *L. lanuginosus* (9 lineages) were very similar to distances between morphospecies, reaching 0.17 in both COI and COII (Table 3.1). Sequences of both 28S rDNA and EF1-α varied within the two morphospecies, showing their limited value as markers for morphospecies in the *L. lanuginosus* species group (Table 3.1).

In general, mean intra-lineage K2P and uncorrected p-distances were smaller compared to inter-lineage distances for both COI and COII in the *L. lanuginosus* species group (supplementary Table S3.4). The mean intra-lineage K2P distances were below 0.032 for both COI and COII, with the largest distances (0.053 for COI and 0.059 for COII) within L13. The mean inter-lineage distances were 0.115–0.252 for COI and 0.128–0.295 for COII, with the lowest distances between L13 and L18.

The nuclear markers varied less than the mitochondrial markers both within and between lineages. Elongation factor 1-α was more efficient than 28S rDNA in reflecting the lineages delimited by COI and COII. The mean intra-lineage K2P and uncorrected p-distances of 28S rDNA and EF1-α were zero or below 0.015 in all 19 lineages (supplementary Table S3.4). Nucleotide sequences of 28S rDNA were identical in L14 and L18 and in L5–8, L10–11 and L14, and sequences of EF1-α were identical in

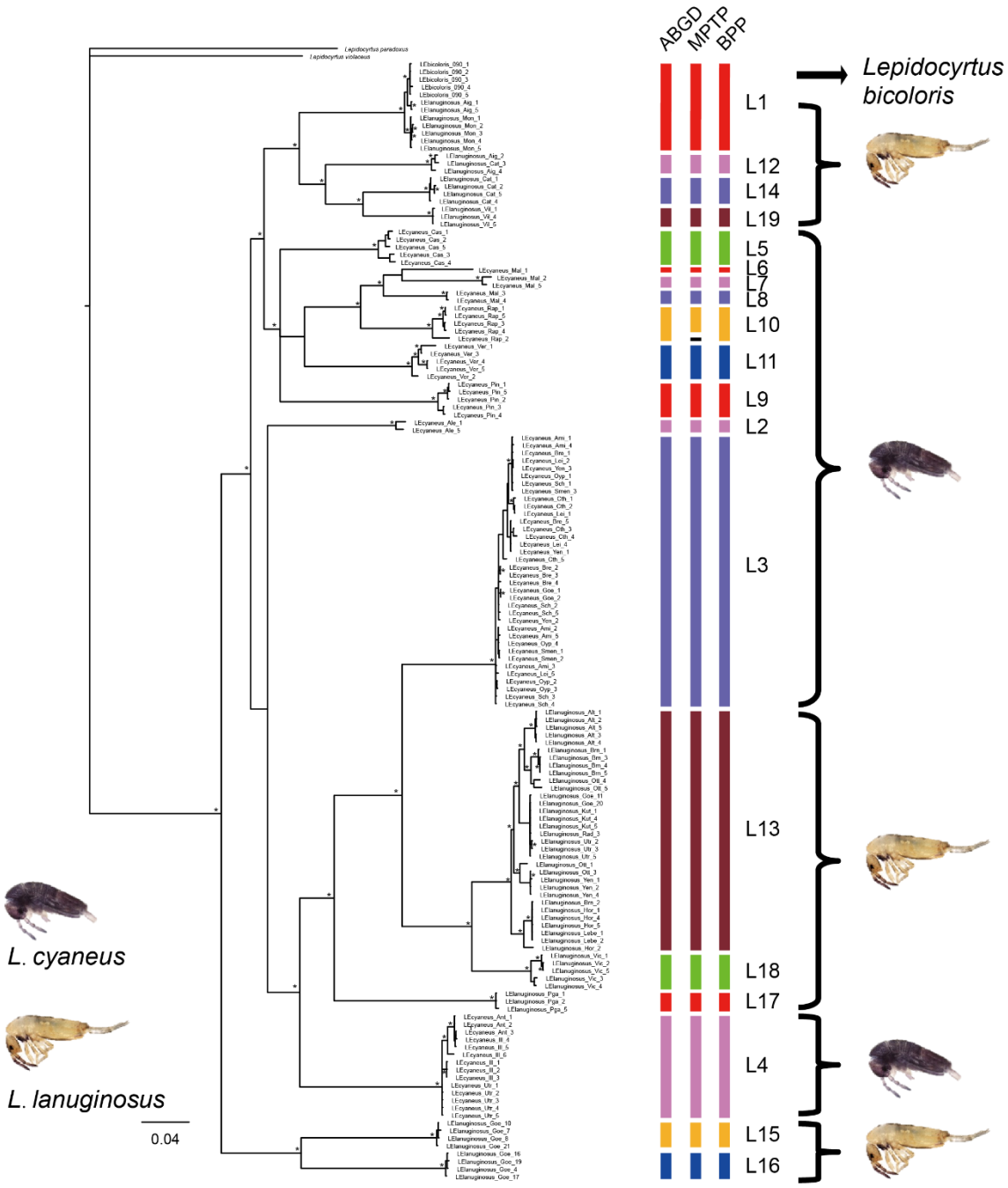


Figure 3.1 Bayesian consensus tree of species based on concatenated data set of COI, COII, 28S rDNA D1-2 and EF1-α and the species delimitation analyses. Asterisks represent the node support with maximum bootstrap values greater than 85 and Bayesian posterior probabilities greater than 0.95. Colored bars represent hypothesized species groupings based on corresponding delimitation analyses based on concatenated data set of COI and COII.

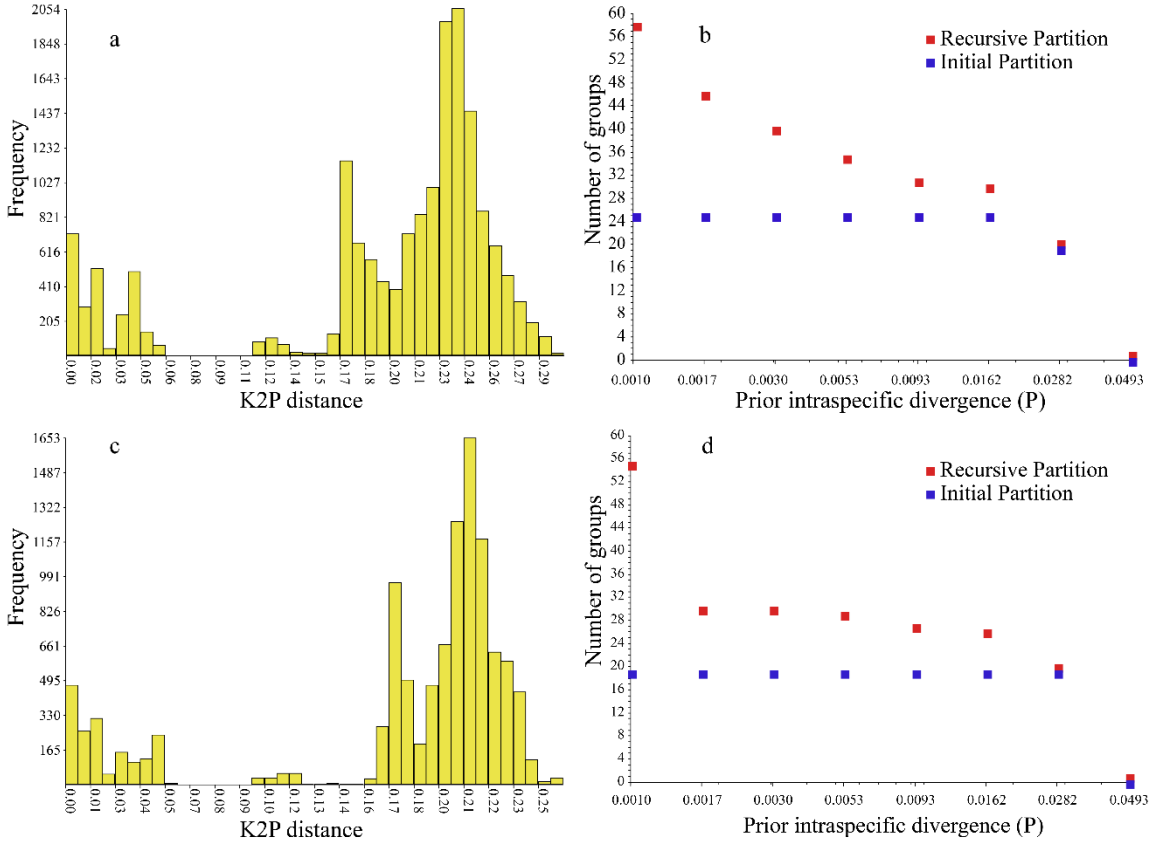


Figure 3.2 Automatic Barcode Gap Discovery species delimitation based on COI: Frequency histogram of K2P pairwise distances (a). Partitions under different prior intra-specific divergences (b). ABGD species delimitation based on COII: Frequency histogram of K2P pairwise distances (c). Partitions under different prior intra-specific divergences (d).

Table 3.1 Mean K2P (upper rows) and uncorrected p-distances (lower rows in gray background) of four genetic markers between different lineages within and between *Lepidocyrtus cyaneus* and *L. lanuginosus* species complexes. Data represent the average ± stand errors.

	COI	COII	28S D1–2	EF1-α
Within the *L. cyaneus* species complex (10 lineages)	0.197±0.022	0.211±0.032	0.004±0.004	0.026±0.01
	0.171±0.017	0.174±0.023	0.004±0.004	0.025±0.01
Within the *L. lanuginosus* species complex (9 lineages)	0.205±0.023	0.244±0.039	0.016±0.01	0.029±0.014
	0.177±0.018	0.198±0.028	0.015±0.01	0.028±0.013
Between *L. cyaneus* and *L. lanuginosus* species complexes	0.209±0.017	0.238±0.024	0.01±0.009	0.028±0.011
	0.181±0.013	0.193±0.016	0.01±0.009	0.027±0.01

L7 and L8, and in L14 and L18 (supplementary Table S3.4). Mean K2P distances between lineages were 0.019 with a range of 0–0.05 for EF1-α, higher than that in 28S rDNA (mean inter-lineage K2P distances 0.009 with a range of 0–0.032).

3.3.3 Phylogenetic analysis

Topologies of ML and BI trees based on the concatenated dataset of the four markers were similar. The majority of lineages of the two morphospecies clustered in separate clades with high support (Bayesian posterior probabilities, pp = 1, Figure 3.1). The phylogenetic trees indicated that both *L. cyaneus* and *L. lanuginosus* are polyphyletic and *L. bicoloris* was grouped with *L. lanuginosus* L1 with strong support (pp = 0.96, Figure 3.1). Overall, the three morphospecies *L. bicoloris*, *L. cyaneus* and *L. lanuginosus* formed a monophyletic clade in the *L. lanuginosus* species group with strong support (pp = 1, Figure 3.1).

3.4 Discussion

3.4.1 High cryptic diversity of the *Lepidocyrtus lanuginosus* species group

Approaches based on genetic distances and evolutionary models showed that both species, *L. cyaneus* and *L. lanuginosus*, comprised multiple genetic lineages / cryptic species. The similar K2P distances of both mitochondrial and nuclear markers within and between these two morphologically defined species suggest that our first hypothesis has to be rejected. ABGD delimited a distance gap of 0.055–0.095 based on COI in the *L. lanuginosus* species group which is similar to other Collembola species groups such as *Protaphorura* (Sun et al., 2017) and *Coecobrya* (F. Zhang et al., 2018). The intra-lineage distances in the *L. lanuginosus* species group of about 0.05 were significantly smaller than the inter-lineage distances (around 0.2). This is also supported by other studies that suggested the existence of distinct lineages / cryptic species in Collembola (Anslan and Tedersoo, 2015; Katz et al., 2015a; Nilsai et al., 2017; F. Zhang et al., 2018). Numbers of lineages delimited by ABGD based on COI and COII were congruent and resembled the results of the BPP analysis; while the PTP analyses based on the concatenated dataset of COI and COII distinguished one more lineage. This incongruence was due to the five individuals of L10 from locality Rapallo in Italy which were assigned to two lineages (supplementary Table S3.1 and Figure 3.1). However, we considered them as one lineage since the K2P distance of COI and COII between these two lineages were 0.04, i.e. lower than the boundary of the detected gap.

Results of the nuclear markers differed in some respect from those of the mitochondrial genes. Elongation factor 1-α confirmed the presence of 17 lineages with nearly no intra-lineage variation (except in L1 and L5) and supports the efficiency of EF1-α in delimiting lineages of *L. lanuginosus* from Southern Europe (Cicconardi et al., 2010). Sequences of 28S rDNA D1–2 failed to delimit all lineages within the *L. lanuginosus* species group as sequences from several lineages were identical: inter-lineage K2P distances were zero in 28S rDNA D1–2 but varied between 0.13 and 0.23 in COI and COII. The D1–2 fragment of the 28S rDNA gene proved to efficiently reflect the phylogeny of Collembola

(D'Haese, 2002) and also has been used for species identification (Sonnenberg et al., 2007). Anslan and Tedersoo (2015) found no intraspecific variation, whereas the interspecific variability ranged between 0.007 and 0.287 among 33 species suggesting the D1–2 fragment of the 28S rDNA gene may be a good Collembola species marker. However, in our study, 28S rDNA did not separate closely related lineages in the *L. lanuginosus* species group as delimited by mitochondrial genes, which is in agreement with earlier studies (Porco et al., 2012a; Zhang et al., 2014) and demonstrates that its validity as Collembola species marker likely is not universal across all taxa. Due to the low variation in 28S rDNA as compared to COI and COII, 28S rDNA is likely to underestimate the true species diversity in some groups. K2P distances of EF1-α were approximately twice as high as those in 28S rDNA among lineages of the *L. lanuginosus* species group and therefore outperformed the latter as Collembola species marker.

3.4.2 Phylogenetic analysis

The molecular phylogeny based on concatenated data did not support monophyly of *L. cyaneus* or *L. lanuginosus*, thus rejecting the second hypothesis that body color is a good species marker. According to the phylogenetic tree, lineages of the *L. lanuginosus* species group split irrespective of body coloration, suggesting multiple losses or gains of body coloration. Body color is one of the most important characters used in separating species in the genus *Lepidocyrtus* (Soto-Adames, 2002), and has been shown great to be of significant phylogenetic and taxonomic value in *Entomobrya* species occurring in China (Ding et al., 2018). However, results of the present study suggest that body color is insufficient to delineate species in the *L. lanuginosus* species group, supporting that body color pattern variation sometimes are uncorrelated with genetic divergence as demonstrated in North American *Entomobrya* species (Katz et al., 2015b). The value of body color as a species marker in Collembola depends on region, species group and possibly sampling size (Ding et al., 2018), but it can be a useful, easily observable, and valid diagnostic tool when combined with other characters.

Molecular data supported the monophyly of the *L. lanuginosus* species group (Mateos et al., 2018). Compared to the restricted sampling of populations and genetic markers in previous studies (Mateos et al., 2018; B. Zhang et al., 2018), we sampled a wider range of populations across Central and Southern Europe, and included a larger number of populations of two species and two additional genetic markers (COI and 28S rDNA), which supports the validity of chaetotaxy as character for delimiting monophyletic clades in the genus *Lepidocyrtus* (Mateos et al., 2018). However, more effort is needed to distinguish lineages / cryptic species within the *L. lanuginosus* species group. Zhang et al. (2018a) proposed that *L. cyaneus* forms part of the *L. lanuginosus* species complex in a local study.

The larger dataset of this study showed that *L. cyaneus* comprises at least ten lineages. Thus, we suggest that a wide sampling across the range of this species complex is needed for uncovering lineages / cryptic species and to understand their relationships.

3.4.3 DNA-based approaches help discovering species diversity

Lepidocyrtus lanuginosus L1 comprised specimens of the morphospecies *L. bicoloris* and *L. lanuginosus*. This suggests that *L. bicoloris* is a polymorphic species including blue colored and uncolored forms. As the only morphological difference between *L. bicoloris* and *L. lanuginosus* is the body color pattern (blue pigment on th.II to abd.II in *L. bicoloris*, without pigment in *L. lanuginosus*; Mateos 2012), uncolored specimens of *L. bicoloris* in the clade *L. lanuginosus* L1 were misidentified as *L. lanuginosus*. These two morphospecies could only be differentiated by molecular analyses. Molecular data combined with morphological characters may help strengthening species hypotheses (De Queiroz, 2007; Samadi and Barberousse, 2006). Integrative taxonomy and more detailed analyses of morphological characters are needed to substantiate the proposition of the existence of nineteen (or more) cryptic species in the *L. lanuginosus* species group. Genetic information, chaetotaxy and other morphological characters have been combined to delimit species and cryptic Collembola species (Felderhoff et al., 2010; Katz et al., 2015b; Schneider and D'Haese, 2013; Sun et al., 2017; Yu et al., 2017; F. Zhang et al., 2018).

DNA-based approaches used in the present study clearly separated lineages of Central Europe from those in Southern Europe. These lineages need to be inspected for subtle morphological differences. For example, the typical labial seta M_1 has been proposed to be absent in specimens of *L. cyaneus* and *L. lanuginosus* from Fennoscandia and Denmark (Fjellberg, 2007), whereas Mateos (2012) found it to be typically present in adult specimens from Southern Europe. Mateos (2012) also found morphological characters, i.e. labium chaetotaxy, in specimens of *L. lanuginosus* from the Northeastern Iberian Peninsula to vary between populations. Large genetic variation among lineages of the *L. lanuginosus* species group revealed by our study suggests that there might be differences in chaetotaxy among the different lineages. The results of our study highlight the need for additional ecological, evolutionary, and taxonomic research of species complexes in Collembola (and other soil invertebrates). Importantly, cryptic species of Collembola may vary in the colonization of different habitats and in species-environment interrelationships. As shown in this study, *L. cyaneus* L5-11 mainly collected from grasslands surrounded by forests in Southern Europe, while the genetically very distant *L. cyaneus* L3 mainly from open grassland and arable land in Central Europe (Fig. 1, supplementary Table S3.1). Zhang *et al.* (2018a) found that different lineages / cryptic species of *L. lanuginosus*

preferentially colonize certain habitats, suggesting that DNA-based approaches are useful to uncover species- and / or lineages-environment interrelationships in soil mesofauna, especially when samples are collected from a large geographical area or different habitats.

The genetically diverse lineages detected by Cicconardi *et al.* (2010), Zhang *et al.* (2018a), and in this study, and the morphological variability of *L. lanuginosus* documented in Mateos (2012), indicates that sampling additional habitats would reveal even more lineages / cryptic species in the *L. lanuginosus* species group. For gaining a coherent picture on the distribution, genetic structure and species-environment interrelationships of lineages on *L. lanuginosus* a wide sampling is indispensable. The number of published European *Lepidocyrtus* species has increased in recent years to 33 (Mateos 2008a, b, 2011, 2012; Traser and Dányi 2008; Mateos and Petersen 2012; Winkler and Traser 2012; Winkler 2016, 2017). DNA-based approaches may accelerate the discovery of new species and provide guidelines for identification of morphological characters of genetically distinct lineages thereby helping to document the full species diversity of European *Lepidocyrtus* and other Collembola species.

3.5 Conclusions

DNA-based approaches revealed distinct MOTUs / lineages in the *L. lanuginosus* species group, suggesting the existence of cryptic species in both *L. cyaneus* and *L. lanuginosus*. Phylogenetic analyses based on nuclear and mitochondrial markers indicated that both *L. cyaneus* and *L. lanuginosus* are polyphyletic. The results suggest that body color is insufficient to distinguish species in the *L. lanuginosus* species group. Further, our results indicate that COII is as efficient as COI in delimiting lineages of the *L. lanuginosus* species group and may be used as alternative DNA barcoding marker in Collembola. Elongation factor 1-α had a higher resolution than 28S rDNA D1–2 in distinguishing closely related genetic lineages, suggesting that it is a more appropriate species marker than 28S rDNA D1–2 for the *L. lanuginosus* species group but less powerful than COI and COII. Overall, this study provides guidelines for understanding and documenting the complex structure of the *L. lanuginosus* species group by integrating genetic and morphological characters. We recommend that phylogeographic studies on closely related species based on a wide range of sampling are needed before analyzing their phylogenetic relationships.

Acknowledgements

We thank Andreas Klein and Bastian Heimburger for the help in sample collection. The study was supported by the German Ministry of Education, Science and Technology (BMBF). Bing Zhang was supported by the China Scholarship Council.

References

Anslan, S., Tedersoo, L., 2015. Performance of cytochrome c oxidase subunit I (COI), ribosomal DNA Large Subunit (LSU) and Internal Transcribed Spacer 2 (ITS2) in DNA barcoding of Collembola. Eur. J. Soil Biol. 69, 1–7. doi:10.1016/j.ejsobi.2015.04.001

Bellinger, P.F., Christiansen, K.A., Janssens, F., 1996. Checklist of the Collembola of the World. http://www.collembola.org/ (accessed 6.26.18).

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K., Das, I., 2007. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. 22, 148–155. doi:10.1016/j.tree.2006.11.004

Carstens, B.C., Pelletier, T.A., Reid, N.M., Satler, J.D., 2013. How to fail at species delimitation. Mol. Ecol. 22, 4369–4383. doi:10.1111/mec.12413

Cicconardi, F., Fanciulli, P.P., Emerson, B.C., 2013. Collembola, the biological species concept and the underestimation of global species richness. Mol. Ecol. 22, 5382–5396. doi:10.1111/mec.12472

Cicconardi, F., Nardi, F., Emerson, B.C., Frati, F., Fanciulli, P.P., 2010. Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. Mol. Ecol. 19, 386–400. doi:10.1111/j.1365-294X.2009.04457.x

D'Haese, C.A., 2002. Were the first springtails semi-aquatic? A phylogenetic approach by means of 28S rDNA and optimization alignment. Proc. R. Soc. B Biol. Sci. 269, 1143–1151. doi:10.1098/rspb.2002.1981

De Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886. doi:10.1080/10635150701701083

Ding, Y., Yu, D., Guo, W., Li, J., Zhang, F., 2018. Molecular phylogeny of *Entomobrya* (Collembola: Entomobryidae) from China: Color pattern groups and multiple origins. Insect Sci. 1–11. doi:10.1111/1744-7917.12559

Felderhoff, K.L., Bernard, E.C., Moulton, J.K., 2010. Survey of *Pogonognathellus* Börner (Collembola: Tomoceridae) in the Southern Appalachians based on morphological and molecular data. Ann. Entomol. Soc. Am. 103, 472–491. doi:10.1603/AN09105

Feulner, P.G.D., Kirschbaum, F., Schugardt, C., Ketmaier, V., Tiedemann, R., 2006. Electrophysiological and molecular genetic evidence for sympatrically occuring cryptic species in African weakly electric fishes (Teleostei: Mormyridae: *Campylomormyrus*). Mol. Phylogenet. Evol. 39, 198–208. doi:10.1016/j.ympev.2005.09.008

Fjellberg, A., 2007. The Collembola of Fennoscandia and Denmark, Part II: Entomobryomorpha and Symphypleona, Brill. Brill, Boston. doi:10.1163/ej.9789004157705.i-265

Grundt, H.H., Kjolner, S., Borgen, L., Rieseberg, L.H., Brochmann, C., 2006. High biological species diversity in the arctic flora. Proc. Natl. Acad. Sci. 103, 972–975. doi:10.1073/pnas.0510270103

Hailer, F., Kutschera, V.E., Hallström, B.M., Klassert, D., Fain, S.R., Leonard, J.A., Arnason, U., Janke, A., 2012. Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. Science (80-.). 336, 344–347. doi:10.1126/science.1216424

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98. doi:citeulike-article-id:691774

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., Hallwachs, W., 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Natl. Acad. Sci. 101, 14812–14817. doi:10.1073/pnas.0406166101

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570–580. doi:10.1093/molbev/msp274

Hopkin, S.P., 2007. A Key to the Springtails (Collembola) of Britain and Ireland, Field Studies Council (AIDGAP Project). FSC Publicatio, Shrewsbury.

Hopkin, S.P., 1997. Biology of the springtails: (Insecta: Collembola). Oxford University Press, Oxford, UK.

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. doi:10.1093/bioinformatics/btx025

Katz, A.D., Giordano, R., Soto-Adames, F., 2015a. Taxonomic review and phylogenetic analysis of fifteen North American *Entomobrya* (Collembola, Entomobryidae), including four new species. Zookeys 2015, 1–75. doi:10.3897/zookeys.525.6020

Katz, A.D., Giordano, R., Soto-Adames, F.N., 2015b. Operational criteria for cryptic species delimitation when evidence is limited, as exemplified by North American *Entomobrya* (Collembola: Entomobryidae). Zool. J. Linn. Soc. 173, 818–840. doi:10.1111/zoj.12220

Kempson, D., Lloyd, M., Ghelardi, R., 1963. A new extractor for woodland litter. Pedobiologia. 3, 1–30.

Knowles, L.L., Carstens, B.C., 2007. Delimiting species without monophyletic gene trees. Syst. Biol. 56, 887–895. doi:10.1080/10635150701701091

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34, 772–773. doi:10.1093/molbev/msw260

Leaché, A.D., Fujita, M.K., 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). Proc. R. Soc. B Biol. Sci. 277, 3071–3077. doi:10.1098/rspb.2010.0662

Mateos, E., 2012. The European *Lepidocyrtus lanuginosus* group (Collembola: Entomobryidae), definition and description of a new species from Spain. Zootaxa 3570, 69–81.

Mateos, E., 2011. New *Lepidocyrtus* Bourlet, 1839 taxa from Greece (Collembola: Entomobryidae). Zootaxa 3108, 25–40.

Mateos, E., 2008a. The European *Lepidocyrtus* Bourlet, 1839 (Collembola: Entomobryidae). Zootaxa 1769, 35–59.

Mateos, E., 2008b. Definition of *Lepidocyrtus lusitanicus* Gama, 1964 species-complex (Collembola, Entomobryidae), with description of new species and color forms from the Iberian Peninsula. Zootaxa 1917, 38–54. doi:10.5281/zenodo.184638

Mateos, E., Escuer, P., Buşmachiu, G., Riutort, M., Álvarez-Presas, M., 2018. Untangling *Lepidocyrtus* (Collembola, Entomobryidae): new molecular data shed light on the relationships of the European groups. Invertebr. Syst. 32, 639–651. doi:10.1071/IS17056

Mateos, E., Petersen, H., 2012. Definition of the European *Lepidocyrtus curvicollis* group (Collembola: Entomobryidae) with description of a new species from Sardinia (Italy). Zootaxa 3273, 51–62.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, in: 2010 Gateway Computing Environments Workshop (GCE). IEEE, pp. 1–8. doi:10.1109/GCE.2010.5676129

Nilsai, A., Jantarit, S., Satasook, C., Zhang, F., 2017. Three new species of coecobrya (Collembola: Entomobryidae) from caves in the Thai peninsula. Zootaxa 4286, 187–202. doi:10.11646/zootaxa.4286.2.3

Porco, D., Bedos, A., Greenslade, P., Janion, C., Skarzyński, D., Stevens, M.I., Jansen van Vuuren, B., Deharveng, L., 2012a. Challenging species delimitation in Collembola: cryptic diversity among common springtails unveiled by DNA barcoding. Invertebr. Syst. 26, 470–477. doi:10.1071/IS12026

Porco, D., Potapov, M., Bedos, A., Busmachiu, G., Weiner, W.M., Hamra-Kroua, S., Deharveng, L., 2012b. Cryptic diversity in the ubiquist species *Parisotoma notabilis* (Collembola, Isotomidae): a long-used chimeric species? PLoS One 7, e46056. doi:10.1371/journal.pone.0046056

Puillandre, N., Lambert, A., Brouillet, S., Achaz, G., 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol. Ecol. 21, 1864–1877. doi:10.1111/j.1365-294X.2011.05239.x

Rambaut, A., Drummond, A.J., M., S., 2013. Tracer v1.6 [WWW Document]. URL http://beast.bio.ed.ac.uk/Tracer (accessed 8.30.18).

Rannala, B., Yang, Z., 2013. Improved reversible jump algorithms for Bayesian species delimitation. Genetics 194, 245–253. doi:10.1534/genetics.112.149039

Rannala, B., Yang, Z., 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164, 1645–1656.

Ronquist, F., Teslenko, M., Van Der Mark, 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. Syst. Biol. 61, 539–542. doi:10.1093/sysbio/sys029

Rougerie, R., Decaëns, T., Deharveng, L., Porco, D., James, S.W., Chang, C.H., Richard, B., Potapov, M., Suhardjono, Y., Hebert, P.D.N., 2009. DNA barcodes for soil animal taxonomy. Pesqui. Agropecuária Bras. 44, 789–802. doi:10.1590/S0100-204X2009000800002

Rusek, J., 2002. Taxonomy of Collembola at the beginning of the new millennium. Pedobiologia. 46, 215–224. doi:10.1078/0031-4056-00128

Salmon, S., Ponge, J.F., Gachet, S., Deharveng, L., Lefebvre, N., Delabrosse, F., 2014. Linking species, traits and habitat characteristics of Collembola at European scale. Soil Biol. Biochem. 75, 73–85. doi:10.1016/j.soilbio.2014.04.002

Samadi, S., Barberousse, A., 2006. The tree, the network, and the species. Biol. J. Linn. Soc. 89, 509–521. doi:10.1111/j.1095-8312.2006.00689.x

Schneider, C., D'Haese, C.A., 2013. Morphological and molecular insights on *Megalothorax*: The largest Neelipleona genus revisited (Collembola). Invertebr. Syst. 27, 317–364. doi:10.1071/IS13002

Sonnenberg, R., Nolte, A.W., Tautz, D., 2007. An evaluation of LSU rDNA D1-D2 sequences for their use in species identification. Front. Zool. 4, 6. doi:10.1186/1742-9994-4-6

Soto-Adames, F.N., 2002. Molecular phylogeny of the Puerto Rican *Lepidocyrtus* and *Pseudosinella* (Hexapoda: Collembola), a validation of Yoshii's "color pattern species." Mol. Phylogenet. Evol. 25, 27–42. doi:10.1016/S1055-7903(02)00250-6

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. doi:10.1093/bioinformatics/btu033

Struck, T.H., Feder, J.L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V.I., Kistenich, S., Larsson, K.H., Liow, L.H., Nowak, M.D., Stedje, B., Bachmann, L., Dimitrov, D., 2017. Finding evolutionary processes hidden in cryptic species. Trends Ecol. Evol. 33, 153–163. doi:10.1016/j.tree.2017.11.007

Sun, X., Zhang, F., Ding, Y., Davies, T.W., Li, Y., Wu, D., 2017. Delimiting species of *Protaphorura* (Collembola: Onychiuridae): integrative evidence based on morphology, DNA sequences and geography. Sci. Rep. 7, 8261. doi:10.1038/s41598-017-08381-4

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. Mol. Biol. Evol. 28, 2731–2739. doi:10.1093/molbev/msr121

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680. doi:10.1093/nar/22.22.4673

Traser, G., Dányi, L., 2008. *Lepidocyrtus mariani* sp. n., a new springtail species from Hungary (Collembola: Entomobryidae). Opusc. Entomol. 39, 91–98.

Vaidya, G., Lohman, D.J., Meier, R., 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27, 171–180. doi:10.1111/j.1096-0031.2010.00329.x

von Saltzwedel, H., Scheu, S., Schaefer, I., 2017. Genetic structure and distribution of *Parisotoma notabilis* (Collembola) in Europe: Cryptic diversity, split of lineages and colonization patterns. PLoS One 12, e0170909. doi:10.1371/journal.pone.0170909

Wiens, J.J., 2007. Species delimitation: new approaches for discovering diversity. Syst. Biol. 56, 875–878. doi:10.1080/10635150701748506

Winkler, D., 2017. New *Lepidocyrtus* Bourlet, 1839 from riverine woodland in Hungary (Collembola, Entomobryidae). Zootaxa 4250, 529–540. doi:10.11646/zootaxa.4250.6.2

Winkler, D., 2016. A new species of *Lepidocyrtus* (Collembola, Entomobryidae) from the Börzsöny Mountains, Hungary. Zootaxa 4150, 388–400. doi:10.11646/zootaxa.4150.4.2

Winkler, D., Traser, G.N., 2012. Explanation of the European *Lepidocyrtus pallidus-serbicus* group (Collembola, Entomobryidae), with description of new species from Hungary. Zootaxa 3394, 35–47.

Yang, Z., Rannala, B., 2017. Bayesian species identification under the multispecies coalescent provides significant improvements to DNA barcoding analyses. Mol. Ecol. 26, 3028–3036. doi:10.1111/mec.14093

Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. 107, 9264–9269. doi:10.1073/pnas.0913022107

Yu, D., Ding, Y., Ma, Y., 2017. Revision of *Tomocerus similis* Chen & Ma, with discussion of the *kinoshitai* complex and the distal tibiotarsal chaetae in Tomocerinae (Collembola, Tomoceridae). Zootaxa 4268, 395–410. doi:10.11646/zootaxa.4268.3.5

Zhang, B., Chen, T.W., Mateos, E., Scheu, S., Schaefer, I., 2018. Cryptic species in *Lepidocyrtus lanuginosus* (Collembola: Entomobryidae) are sorted by habitat type. Pedobiologia. 68, 12–19. doi:10.1016/j.pedobi.2018.03.001

Zhang, C., Rannala, B., Yang, Z., 2012. Robustness of compound dirichlet priors for bayesian inference of branch lengths. Syst. Biol. 61, 779–784. doi:10.1093/sysbio/sys030

Zhang, F., Jantarit, S., Nilsai, A., Stevens, M.I., Ding, Y., Satasook, C., 2018. Species delimitation in the morphologically conserved *Coecobrya* (Collembola: Entomobryidae): A case study integrating morphology and molecular traits to advance current taxonomy. Zool. Scr. 47, 342–356. doi:10.1111/zsc.12279

Zhang, F., Yu, D., Luo, Y., Ho, S.Y.W., Wang, B., Zhu, C., 2014. Cryptic diversity, diversification and vicariance in two species complexes of *Tomocerus* (Collembola, Tomoceridae) from China. Zool. Scr. 43, 393–404. doi:10.1111/zsc.12056

Zhang, J., Kapli, P., Pavlidis, P., Stamatakis, A., 2013. A general species delimitation method with applications to phylogenetic placements. Bioinformatics 29, 2869–2876. doi:10.1093/bioinformatics/btt499

Supplementary Materials

Supplementary Table S3.1 Sampling data of Collembola specimens used for DNA barcoding. 28SD1–2, COII, COI and elongation factor 1-α (EF 1-α) accessions represents the registration numbers of sequences in GenBank. Loc. Abb.: Locality abbreviations.

Species	Individual name	Locality	Loc. Abb.	Country	Habitat	Position	Collecting date	28SD1_2	COII	COI	EF 1-α
Lepidocyrtus bicoloris	LEbicoloris_090_1	Tagamanent	Tag	Spain	Grassland surround by forest	41°45'3.68"N, 2°18'20.50"E	15/10/2011	MH570534	MH570690		MH570984
	LEbicoloris_090_2							MH570535	MH570691		MH570985
	LEbicoloris_090_3							MH570536	MH570692	MH570854	MH570986
	LEbicoloris_090_4							MH570537	MH570693		
	LEbicoloris_090_5							MH570538	MH570694	MH570855	
Lepidocyrtus cyaneus	LEcyaneus_Ale_1	Ax-les-Thermes	Ale	France	Grassland surround by forest	42°44'33.41"N, 1°46'59.04"E	11/05/2015	MH570539	MH570695	MH570856	MH570987
	LEcyaneus_Ale_2							MH570540	MH570696	MH570857	MH570988
	LEcyaneus_Ale_3							MH570541	MH570697	MH570858	MH570989
	LEcyaneus_Ale_4							MH570542	MH570698	MH570859	MH570990
	LEcyaneus_Ale_5							MH570543	MH570699	MH570860	MH570991
	LEcyaneus_Ami_1	Amiens	Ami	France	Grassland next to forest	42°44'33.41"N, 1°46'59.04"E	11/05/2015	MH570544	MH570700		
	LEcyaneus_Ami_2							MH570545	MH570701	MH570861	MH570992
	LEcyaneus_Ami_3							MH570546	MH570702	MH570862	MH570993
	LEcyaneus_Ami_4							MH570547	MH570703		MH570994
	LEcyaneus_Ami_5							MH570548	MH570704	MH570863	MH570995
	LEcyaneus_Ant_1	Antwerpen	Ant	Belgium	Grassland-arable land	49°50'33.9"N, 2°31'35.7"E	18/04/2016	MH570549	MH570705		MH570996
	LEcyaneus_Ant_2							MH570550	MH570706	MH570864	MH570997
	LEcyaneus_Ant_3							MH570551	MH570707	MH570865	MH570998
	LEcyaneus_Ant_4							MH570552	MH570708	MH570866	MH570999
	LEcyaneus_Ant_5							MH570553	MH570709	MH570867	MH571000
	LEcyaneus_Bre_1	Bremen North	Bre	Germany		53°14'36.1"N, 8°39'20.1"E	19/04/2016	MH570554	MH570710	MH570868	MH571001
	LEcyaneus_Bre_2							MH570555	MH570711	MH570869	MH571002

Species	Individual name	Locality	Loc. Abb.	Country	Habitat	Position	Collecting date	28SD1_2	COII	COI	EF 1-α
	LEcyaneus_Bre_3				Grassland next to forest				MH570712	MH570870	MH571003
	LEcyaneus_Bre_4							MH570556	MH570713	MH570871	MH571004
	LEcyaneus_Bre_5								MH570714	MH570872	MH571005
	LEcyaneus_Cas_1	Cassine	Cas	Italy	Grassland surround by forest	44°46'8.09"N, 8°29'28.64"E	07/05/2015	MH570557	MH570715	MH570873	MH571006
	LEcyaneus_Cas_2							MH570558	MH570716	MH570874	MH571007
	LEcyaneus_Cas_3							MH570559	MH570717	MH570875	MH571008
	LEcyaneus_Cas_4							MH570560	MH570718		
	LEcyaneus_Cas_5							MH570561	MH570719	MH570876	MH571009
	LEcyaneus_Cth_1	Chateau-Thierry	Cth	France	Grassland-arable land	49° 5'45.17"N, 3°26'12.35"E	13/05/2015	MH570562	MH570720		MH571010
	LEcyaneus_Cth_2							MH570563	MH570721	MH570877	MH571011
	LEcyaneus_Cth_3							MH570564	MH570722		MH571012
	LEcyaneus_Cth_4							MH570565	MH570723	MH570878	MH571013
	LEcyaneus_Cth_5							MH570566	MH570724	MH570879	MH571014
	LEcyaneus_Goe_1	Ossenfeld grassland	Ossg	Germany	Grassland next to forest	51°32'50.3"N, 9°47'50.4"E	17/04/2014	MH160159	MH144024	MH143939	MH178012
	LEcyaneus_Goe_2							MH160160	MH144025	MH143940	MH178013
	LEcyaneus_Goe_3	Waake arable land	Waaa	Germany	Arable land	51°33'47.1"N, 10°03'30.4"E	19/06/2014	MH160161	MH144026	MH143941	MH178014
	LEcyaneus_Goe_4							MH160162	MH144027	MH143942	MH178015
	LEcyaneus_Goe_5	Ossenfeld arable land	Ossa	Germany	Arable land	51°32'52.4"N, 9°47'52.9"E	17/04/2014	MH160158	MH144023	MH143938	MH178011
	LEcyaneus_Ill_1	Illertissen	Ill	Germany	Grassland next to forest	48°13'41.43"N, 10° 7'1.12"E	03/05/2015	MH570567	MH570725	MH570880	
	LEcyaneus_Ill_2							MH570568	MH570726	MH570881	MH571015
	LEcyaneus_Ill_3							MH570569	MH570727	MH570882	MH571016
	LEcyaneus_Ill_4							MH570570	MH570728	MH570883	MH571017
	LEcyaneus_Ill_5							MH570571	MH570729	MH570884	
	LEcyaneus_Ill_6							MH570572	MH570730	MH570885	
	LEcyaneus_Jen_1	Jena	Jen	Germany			12/05/2016	MH570573	MH570731	MH570886	MH571018

Species	Individual name	Locality	Loc. Abb.	Country	Habitat	Position	Collecting date	28SD1_2	COII	COI	EF 1-α
	LEcyaneus_Jen_2							MH570574	MH570732	MH570887	
	LEcyaneus_Jen_3				Grassland-arable land	50°57'0.96"N, 11°37'16.43"E		MH570575	MH570733	MH570888	MH571019
	LEcyaneus_Jen_4							MH570576	MH570734	MH570889	MH571020
	LEcyaneus_Jen_5							MH570577	MH570735	MH570890	MH571021
	LEcyaneus_Lei_1	Leipzig	Lei	Germany	Grassland surround by forest	51°19'33.78"N, 12°21'21.05"E	25/04/2016	MH570578	MH570736	MH570891	MH571022
	LEcyaneus_Lei_2							MH570579	MH570737	MH570892	MH571023
	LEcyaneus_Lei_3							MH570580	MH570738	MH570893	MH571024
	LEcyaneus_Lei_4							MH570581	MH570739	MH570894	MH571025
	LEcyaneus_Lei_5							MH570582	MH570740		
	LEcyaneus_Mal_1	Marseille-Allauch	Mal	France	Grassland surround by forest	43°22'16.57"N, 5°32'9.46"E	09/05/2015	MH570583	MH570741	MH570895	MH571026
	LEcyaneus_Mal_2							MH570584	MH570742	MH570896	MH571027
	LEcyaneus_Mal_3							MH570585	MH570743	MH570897	MH571028
	LEcyaneus_Mal_4							MH570586	MH570744		MH571029
	LEcyaneus_Mal_5							MH570587	MH570745		
	LEcyaneus_Oyp_1	Oye-Plage (Calais)	Oyp	France	Grassland next to forest	50°59'47.5"N, 2°02'31.9"E	18/04/2016	MH570588	MH570746	MH570898	
	LEcyaneus_Oyp_2							MH570589	MH570747	MH570899	
	LEcyaneus_Oyp_3							MH570590	MH570748	MH570900	
	LEcyaneus_Oyp_4							MH570591	MH570749		
	LEcyaneus_Oyp_5							MH570592	MH570750		
	LEcyaneus_Pin_1	"Ordesa y Monte Perdido" National Park, Pineta valley	Pin	Spain	Grassland surround by forest	42°40'35.3"N, 0°05'07.5"E	30/05/2009	MH570593	MH570751	MH570901	MH571030
	LEcyaneus_Pin_2							MH570594	MH570752	MH570902	MH571031
	LEcyaneus_Pin_3							MH570595	MH570753		MH571032
	LEcyaneus_Pin_4							MH570596	MH570754		MH571033
	LEcyaneus_Pin_5							MH570597	MH570755		MH571034
	LEcyaneus_Rap_1	Rapallo	Rap	Italy	Grassland surround by forest	44°22'48.53"N, 9°30'40.63"E	06/05/2015	MH570598	MH570756	MH570903	
	LEcyaneus_Rap_2							MH570599	MH570757	MH570904	MH571035
	LEcyaneus_Rap_3							MH570600	MH570758		

Species	Individual name	Locality	Loc. Abb.	Country	Habitat	Position	Collecting date	28SD1_2	COII	COI	EF 1-α
	LEcyaneus_Rap_4							MH570601	MH570759	MH570905	
	LEcyaneus_Rap_5							MH570602	MH570760	MH570906	MH571036
	LEcyaneus_Sch_1	Schwechat (Wien Airport)	Sch	Austria	Grassland next to forest	48°08'38.0"N, 16°31'32.1"E	23/04/2016	MH570603	MH570761	MH570907	MH571037
	LEcyaneus_Sch_2							MH570604	MH570762	MH570908	MH571038
	LEcyaneus_Sch_3							MH570605	MH570763	MH570909	MH571039
	LEcyaneus_Sch_4							MH570606	MH570764	MH570910	MH571040
	LEcyaneus_Sch_5							MH570607	MH570765	MH570911	MH571041
	LEcyaneus_Smen_1	Sainte-Menehoult	Smen	France	Grassland-arable land	49° 5'46.17"N, 4°55'26.67"E	14/05/2015	MH570608	MH570766	MH570912	MH571042
	LEcyaneus_Smen_2							MH570609	MH570767	MH570913	MH571043
	LEcyaneus_Smen_3							MH570610	MH570768	MH570914	MH571044
	LEcyaneus_Smen_4							MH570611	MH570769	MH570915	MH571045
	LEcyaneus_Smen_5							MH570612	MH570770	MH570916	MH571046
	LEcyaneus_Utr_1	Utrecht	Utr	The Netherlands	Grassland surround by forest	52°08'50.6"N, 5°12'21.9"E	19/04/2016	MH570613	MH570771	MH570917	MH571047
	LEcyaneus_Utr_2							MH570614	MH570772	MH570918	MH571048
	LEcyaneus_Utr_3							MH570615	MH570773	MH570919	MH571049
	LEcyaneus_Utr_4								MH570774		MH571050
	LEcyaneus_Utr_5							MH570616	MH570775	MH570920	MH571051
	LEcyaneus_Ver_1	Vernante	Ver	Italy	Grassland surround by forest	44°12'6.99"N, 7°30'17.58"E	07/05/2015	MH570617	MH570776	MH570921	MH571052
	LEcyaneus_Ver_2							MH570618	MH570777	MH570922	MH571053
	LEcyaneus_Ver_3							MH570619	MH570778	MH570923	
	LEcyaneus_Ver_4								MH570779		MH571054
	LEcyaneus_Ver_5							MH570620	MH570780	MH570924	
Lepidocyrtus lanuginosus	LElanuginosus_Aig_1	"Montseny" mountain (Aiguafreda)	Aig	Spain	Forest	41°46'13.8"N, 2°16'19.9"E	10/02/2007	MH570621	MH570781	MH570925	MH571055
	LElanuginosus_Aig_2							MH570622	MH570782	MH570926	MH571056
	LElanuginosus_Aig_4							MH570623	MH570783	MH570927	MH571057
	LElanuginosus_Aig_5							MH570624	MH570784	MH570928	MH571058
	LElanuginosus_Alt_1	Altmoorhausen	Alt	Germany			18/05/2016	MH570625	MH570785	MH570929	MH571059

Species	Individual name	Locality	Loc. Abb.	Country	Habitat	Position	Collecting date	28SD1_2	COII	COI	EF 1-α
	LElanuginosus_Alt_2							MH570626	MH570786	MH570930	MH571060
	LElanuginosus_Alt_3				Grassland-arable land	53° 3'59.74"N, 8°21'1.83"E		MH570627	MH570787	MH570931	MH571061
	LElanuginosus_Alt_4							MH570628	MH570788	MH570932	MH571062
	LElanuginosus_Alt_5							MH570629	MH570789	MH570933	MH571063
	LElanuginosus_Brn_1	Brno	Brn	Czech Republic	Grassland next to forest	49°11'32.15"N, 16°25'33.73"E	24/04/2016	MH570630	MH570790		
	LElanuginosus_Brn_2								MH570791		
	LElanuginosus_Brn_3							MH570631	MH570792		
	LElanuginosus_Brn_4								MH570793	MH570934	
	LElanuginosus_Brn_5								MH570794	MH570935	MH571064
	LElanuginosus_Cat_1	"Serra de Catllaràs" mountain	Cat	Spain	Forest	42°13'44.0"N, 1°56'16.8"E	05/04/2008	MH570632	MH570795	MH570936	MH571065
	LElanuginosus_Cat_2							MH570633	MH570796	MH570937	MH571066
	LElanuginosus_Cat_3							MH570634	MH570797	MH570938	MH571067
	LElanuginosus_Cat_4							MH570635	MH570798	MH570939	MH571068
	LElanuginosus_Cat_5							MH570636	MH570799	MH570940	MH571069
	LElanuginosus_Goe_4	Herberhausen forest	Herf	Germany	Forest	51°31'50.6"N, 9°59'26.4"E	19/06/2014	MH160107	MH143972	MH143922	MH178004
	LElanuginosus_Goe_6	Herberhausen grassland	Herg	Germany	Grassland next to forest	51°31'58.6"N, 9°59'33.8"E	19/06/2014	MH160110	MH143975	MH143925	MH177997
	LElanuginosus_Goe_7							MH160111	MH143976	MH143926	MH177998
	LElanuginosus_Goe_8							MH160113	MH143978	MH143923	MH177995
	LElanuginosus_Goe_10							MH160109	MH143974	MH143924	MH177996
	LElanuginosus_Goe_11	Herberhausen arable land	Hera	Germany	Arable field	51°32'02.7"N, 10°00'02.1"E	19/06/2014	MH160100	MH143965	MH143920	MH177993
	LElanuginosus_Goe_12							MH160101	MH143966	MH143921	MH177994
	LElanuginosus_Goe_13							MH160102	MH143967	MH570982	MH571118
	LElanuginosus_Goe_15	Deppoldshausen arable land	Depa	Germany	Arable field	51°32'02.7"N, 10°00'02.1"E	27/03/2014	MH160114	MH143979	MH143927	MH177999
	LElanuginosus_Goe_16	Ossenfeld forest	Ossf	Germany	Forest	51°32'56.3"N, 9°48'01.5"E	14/04/2014	MH160133	MH143998	MH143930	MH178005

Species	Individual name	Locality	Loc. Abb.	Country	Habitat	Position	Collecting date	28SD1_2	COII	COI	EF 1-α
	LElanuginosus_Goe_17	Waake forest	Waaf	Germany	Forest	51°33'34.0"N, 10°04'14.9"E	19/06/2014	MH160141	MH144006	MH143933	MH178006
	LElanuginosus_Goe_18							MH160144	MH144009	MH143934	MH178007
	LElanuginosus_Goe_19							MH160145	MH144010	MH570983	MH571119
	LElanuginosus_Goe_20	Waake arable land	Waaa	Germany	Arable field	51°33'47.1"N, 10°03'30.4"E	19/06/2014	MH160136	MH144001	MH143931	MH178000
	LElanuginosus_Goe_21	Waake arable land	Waaa	Germany	Arable field and grassland	51°33'47.1"N, 10°03'30.4"E	19/06/2014	MH160140	MH144005	MH143932	MH178001
	LElanuginosus_Hor_1	Horka (Jungbunzlau)	Hor	Czech Republic	Grassland-arable land	50°28'40.2"N, 14°58'43.0"E	24/04/2016	MH570637	MH570800	MH570941	MH571070
	LElanuginosus_Hor_2							MH570638	MH570801		MH571071
	LElanuginosus_Hor_3							MH570639	MH570802	MH570942	MH571072
	LElanuginosus_Hor_4							MH570640	MH570803	MH570943	
	LElanuginosus_Hor_5							MH570641	MH570804		MH571073
	LElanuginosus_Jen_1	Jena	Jen	Germany	Grassland-arable land	50°57'0.96"N, 11°37'16.43"E	12/05/2016	MH570642	MH570805		MH571074
	LElanuginosus_Jen_2							MH570643	MH570806		MH571075
	LElanuginosus_Jen_3							MH570644	MH570807	MH570944	MH571076
	LElanuginosus_Jen_4							MH570645		MH570945	MH571077
	LElanuginosus_Kut_1	Kutno	Kut	Poland	Grassland next to forest	52°16'38.8"N, 19°20'54.4"E	21/04/2016	MH570646	MH570808	MH570946	MH571078
	LElanuginosus_Kut_2							MH570647	MH570809	MH570947	MH571079
	LElanuginosus_Kut_3							MH570648	MH570810	MH570948	MH571080
	LElanuginosus_Kut_4								MH570811	MH570949	MH571081
	LElanuginosus_Kut_5							MH570649	MH570812	MH570950	MH571082
	LElanuginosus_Lebe_1	Lebeny west	Lebe	Hungary	Grassland-arable land	47°44'55.8"N, 17°21'08.7"E	23/04/2016		MH570813	MH570951	
	LElanuginosus_Lebe_2								MH570814	MH570952	MH571083
	LElanuginosus_Lebe_3							MH570650	MH570815	MH570953	MH571084
	LElanuginosus_Lebe_4							MH570651	MH570816	MH570954	MH571085
	LElanuginosus_Mon_1		Mon	Spain	Forest		18/04/2007	MH570652	MH570817	MH570955	MH571086

Species	Individual name	Locality	Loc. Abb.	Country	Habitat	Position	Collecting date	28SD1_2	COII	COI	EF 1-α
	LElanuginosus_Mon_2	"Montnegre" mountain				41°39'52.2"N, 2°33'46.0"E		MH570653	MH570818	MH570956	MH571087
	LElanuginosus_Mon_3							MH570654	MH570819	MH570957	MH571088
	LElanuginosus_Mon_4							MH570655	MH570820	MH570958	
	LElanuginosus_Mon_5							MH570656	MH570821	MH570959	MH571089
	LElanuginosus_Niea_10	Nitra East	Niea	Slovakia	Grassland-arable land	48°18'13.9"N, 18°09'11.8"E	23/04/2016	MH570657	MH570822	MH570960	MH571090
	LElanuginosus_Niea_6							MH570658	MH570823	MH570961	MH571091
	LElanuginosus_Niea_7							MH570659	MH570824	MH570962	MH571092
	LElanuginosus_Niea_8							MH570660	MH570825	MH570963	MH571093
	LElanuginosus_Niea_9							MH570661	MH570826	MH570964	MH571094
	LElanuginosus_Ott_1	Otterbach	Ott	Germany	Grassland surround by forest	49°26'31.74"N, 7°55'19.24"E	14/05/2015	MH570662	MH570827	MH570965	MH571095
	LElanuginosus_Ott_2							MH570663	MH570828	MH570966	MH571096
	LElanuginosus_Ott_3							MH570664	MH570829	MH570967	MH571097
	LElanuginosus_Ott_4							MH570665	MH570830		MH571098
	LElanuginosus_Ott_5							MH570666	MH570831	MH570968	MH571099
	LElanuginosus_Pga_1	Park Gardasee	Pga	Italy	Grassland surround by forest	45°42'15.73"N, 10°58'23.94"E	04/05/2015	MH570667	MH570832		
	LElanuginosus_Pga_2							MH570668	MH570833		MH571100
	LElanuginosus_Pga_3							MH570669	MH570834		MH571101
	LElanuginosus_Pga_5							MH570671	MH570835	MH570969	MH571102
	LElanuginosus_Rad_1	Radom	Rad	Poland	Grassland-arable land	51°22'49.0"N, 21°06'57.5"E	22/04/2016	MH570672	MH570836	MH570970	MH571103
	LElanuginosus_Rad_2							MH570673	MH570837	MH570971	MH571104
	LElanuginosus_Rad_3							MH570674	MH570838	MH570972	MH571105
	LElanuginosus_Rad_4							MH570675	MH570839	MH570973	MH571106
	LElanuginosus_Rad_5							MH570676	MH570840	MH570974	MH571107
	LElanuginosus_Utr_2	Utrecht	Utr	The Netherlands	Grassland surround by forest	52°08'50.6"N, 5°12'21.9"E	19/04/2016	MH570677	MH570841		MH571108
	LElanuginosus_Utr_3							MH570678	MH570842		MH571109
	LElanuginosus_Utr_5							MH570679	MH570843	MH570975	MH571110
	LElanuginosus_Vic_1	Vichy	Vic	France			13/05/2015	MH570680	MH570844	MH570976	

Species	Individual name	Locality	Loc. Abb.	Country	Habitat	Position	Collecting date	28SD1_2	COII	COI	EF 1-α
	LElanuginosus_Vic_2							MH570681	MH570845		
	LElanuginosus_Vic_3				Grassland next to forest	46° 6'24.04"N, 3°22'16.84"E		MH570682	MH570846	MH570977	MH571111
	LElanuginosus_Vic_4							MH570683	MH570847	MH570978	MH571112
	LElanuginosus_Vic_5							MH570684	MH570848	MH570979	MH571113
	LElanuginosus_Vil_1	near EDAR east (Viladecavalls)	Vil	Spain	Forest	41°32'49.2"N, 1°57'36.4"E	25/10/2006	MH570685	MH570849		MH571114
	LElanuginosus_Vil_2							MH570686	MH570850		MH571115
	LElanuginosus_Vil_3							MH570687	MH570851		MH571116
	LElanuginosus_Vil_4							MH570688	MH570852	MH570980	MH571117
	LElanuginosus_Vil_5							MH570689	MH570853	MH570981	
Lepidocyrtus paradoxus	LEparadoxus_Mas_1	Masetti	Mas	Italy	Grassland surround by forest	46° 3'23.22"N, 11°15'29.51"E	04/05/2015	MH160172	MH144037	MH143952	MH178032
Lepidocyrtus cf. *violaceus*	LEviolaceus_Pga_1	Park Gardasee	Pga	Italy	Grassland surround by forest	45°42'15.73"N, 10°58'23.94"E	04/05/2015	MH160175	MH144040	MH143955	MH178023

Supplementary Table S3.2 Sequencing primers and fragment length.

Gene	Primer name	Sequence (5'–3')	Length (bp)	Reference
COI	LCO1490	GGTCAACAAATCATAAAGATATTGG	709	(Folmer et al., 1994)
	HCO2198	TAAACTTCAGGGTGACCAAAAAATCA		(Folmer et al., 1994)
	ColFol-for	TTTCAACAAATCATAARGAYATYGG		(Ramirez-Gonzalez et al., 2013)
	ColFol-rev	TAAACTTCNGGRTGNCCAAAAAATCA		(Ramirez-Gonzalez et al., 2013)
COII	tRNA-20-LcuJ	GGTTTAAGAGACCGTGGCTTAC	750	(Cicconardi et al., 2010)
	tRNA-13-LcuN	TCTAACGTGGCAGACTAGTGC		(Cicconardi et al., 2010)
	tRNA-K-LcuJ	GAGCGTATTATAAAGCGGTTTAAG		(Cicconardi et al., 2010)
	tRNA-L-LcuN	CAGACTAGTGCCATGAATTTAAGC		(Cicconardi et al., 2010)
EF 1-α	EFLcuJ	ATGGGGGCAAGATAGCGTCAA	470-580	(Cicconardi et al., 2010)
	EFLcuN	TGAAGGCTGAACGTGAACGTGG		(Cicconardi et al., 2010)
28S_D1–2	D2	ACCACGCATGCWTTAGATTG	~760	(D'Haese, 2002)
	C1'	ACCCGCTGAATTTAAGCAT		(D'Haese, 2002)
	Lepi_28S_nest_R	GGTACAGACAACGTGCTACGG		

References

Cicconardi, F., Nardi, F., Emerson, B. C., Frati, F., and Fanciulli, P. P. (2010). Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. *Molecular Ecology* 19, 386–400. doi:10.1111/j.1365-294X.2009.04457.x

D'Haese, C. A. (2002). Were the first springtails semi-aquatic? A phylogenetic approach by means of 28S rDNA and optimization alignment. *Proceedings of the Royal Society B: Biological Sciences* 269, 1143–1151. doi:10.1098/rspb.2002.1981

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. Mol. Mar. Biol. Biotechnol. 3, 294–299.

Ramirez-Gonzalez, R., Yu, D.W., Bruce, C., Heavens, D., Caccamo, M., Emerson, B.C., 2013. PyroClean: denoising pyrosequences from protein-coding amplicons for the recovery of interspecific and intraspecific genetic variation. PLoS One 8, 1–11. doi:10.1371/journal.pone.0057615

Supplementary Table S3.3 Best-fitting substitution models for each locus (partition) under the BIC criterion in PartitionFinder v2.1.1.

Subset	Best Model	sites	Partition names
1	HKY+I	709	28S
2	SYM+I+G	227	COII_pos1
3	F81+I+G	227	COII_pos2
4	GTR+G	444	COI_pos3, COII_pos3
5	SYM+G	217	COI_pos1
6	F81+I	217	COI_pos2
7	JC+I	140	EF_pos1, EF_pos3
8	K80+G	70	EF_pos2

Supplementary Table S3.4 Genetic K2P distances within and between lineages of the *Lepidocyrtus lanuginosus* species group.

Gene		Within lineages		Between lineages																
		range	mean																	
				L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18
COI																					
	L1	0–0.016	0.008																		
	L2	0–0.03	0.012	0.209																	
	L3	0–0.02	0.011	0.203	0.208																
	L4	0–0.024	0.012	0.192	0.22	0.208															
	L5	0.005–0.035	0.021	0.191	0.22	0.202	0.177														
	L6	/	/	0.207	0.216	0.19	0.231	0.183													
	L7	/	/	0.216	0.199	0.195	0.204	0.197	0.154												
	L8	/	/	0.187	0.202	0.204	0.226	0.191	0.128	0.149											
	L9	0.002	0.002	0.231	0.214	0.206	0.217	0.178	0.227	0.202	0.205										
	L10	0–0.038	0.021	0.208	0.216	0.212	0.213	0.198	0.162	0.17	0.163	0.21									
	L11	0.002–0.041	0.023	0.199	0.232	0.196	0.203	0.185	0.197	0.187	0.181	0.215	0.183								
	L12	0.005–0.016	0.011	0.192	0.221	0.208	0.213	0.218	0.215	0.186	0.199	0.207	0.218	0.203							
	L13	0–0.053	0.03	0.21	0.198	0.173	0.208	0.224	0.195	0.195	0.193	0.226	0.208	0.194	0.209						
	L14	0–0.006	0.004	0.198	0.229	0.211	0.203	0.166	0.176	0.197	0.187	0.221	0.218	0.188	0.185	0.214					
	L15	0–0.005	0.002	0.187	0.225	0.214	0.224	0.227	0.218	0.218	0.223	0.223	0.231	0.216	0.221	0.234	0.213				
	L16	0–0.003	0.002	0.198	0.252	0.228	0.23	0.194	0.219	0.23	0.194	0.232	0.209	0.212	0.207	0.232	0.212	0.219			
	L17	/	/	0.209	0.232	0.222	0.216	0.192	0.229	0.239	0.174	0.215	0.218	0.22	0.237	0.225	0.193	0.229	0.207		
	L18	0.002–0.028	0.019	0.195	0.217	0.179	0.204	0.209	0.212	0.208	0.198	0.218	0.233	0.215	0.221	0.115	0.221	0.225	0.215	0.206	
	L19	0.003	0.003	0.184	0.25	0.201	0.193	0.18	0.199	0.211	0.185	0.188	0.208	0.192	0.195	0.212	0.143	0.195	0.196	0.204	0.213
COII																					
	L1	0–0.01	0.004																		
	L2	0–0.016	0.007	0.225																	

Gene		Within lineages		Between lineages																	
		range	mean																		
	L3	0–0.028	0.014	0.239	0.198																
	L4	0–0.023	0.012	0.254	0.221	0.219															
	L5	0.003–0.035	0.022	0.207	0.188	0.253	0.258														
	L6	0	/	0.214	0.196	0.214	0.225	0.199													
	L7	0.015	0.015	0.237	0.199	0.23	0.262	0.228	0.132												
	L8	0	0	0.189	0.193	0.246	0.25	0.214	0.139	0.152											
	L9	0–0.016	0.01	0.247	0.234	0.231	0.231	0.229	0.203	0.204	0.204										
	L10	0.002–0.04	0.018	0.211	0.206	0.243	0.258	0.228	0.154	0.163	0.157	0.237									
	L11	0–0.045	0.025	0.224	0.201	0.224	0.265	0.22	0.183	0.202	0.18	0.227	0.213								
	L12	0.011–0.013	0.012	0.188	0.217	0.261	0.266	0.196	0.216	0.234	0.219	0.237	0.229	0.226							
	L13	0–0.059	0.032	0.242	0.232	0.185	0.231	0.247	0.23	0.246	0.247	0.226	0.237	0.241	0.253						
	L14	0–0.006	0.004	0.202	0.205	0.24	0.242	0.196	0.214	0.224	0.194	0.208	0.214	0.211	0.19	0.255					
	L15	0–0.006	0.003	0.265	0.23	0.266	0.248	0.248	0.258	0.252	0.257	0.273	0.246	0.253	0.258	0.28	0.282				
	L16	0–0.006	0.004	0.246	0.255	0.273	0.278	0.269	0.295	0.289	0.258	0.276	0.259	0.278	0.279	0.284	0.288	0.237			
	L17	0–0.003	0.002	0.245	0.239	0.26	0.221	0.245	0.246	0.249	0.254	0.241	0.247	0.257	0.238	0.227	0.235	0.244	0.284		
	L18	0–0.015	0.009	0.25	0.25	0.18	0.248	0.275	0.235	0.229	0.263	0.214	0.244	0.234	0.276	0.128	0.256	0.295	0.28	0.226	
	L19	0	0	0.207	0.21	0.255	0.242	0.202	0.228	0.245	0.229	0.204	0.225	0.225	0.182	0.266	0.145	0.268	0.284	0.244	0.262
28S D1–2																					
	L1	0–0.009																			
	L2	0		0.01																	
	L3	0		0.009	0.009																
	L4	0		0.009	0.006	0.006															
	L5	0		0.004	0.006	0.009	0.006														
	L6	0		0.004	0.006	0.009	0.006	0													
	L7	0		0.004	0.006	0.009	0.006	0	0												
	L8	0		0.004	0.006	0.009	0.006	0	0	0											

Gene		Within lineages		Between lineages																	
		range	mean																		
	L9	0		0.006	0.004	0.01	0.007	0.001	0.001	0.001	0.001										
	L10	0		0.004	0.006	0.009	0.006	0	0	0	0	0.001									
	L11	0		0.004	0.006	0.009	0.006	0	0	0	0	0.001	0								
	L12	0		0.007	0.007	0.01	0.007	0.004	0.004	0.004	0.004	0.006	0.004	0.004							
	L13	0		0.009	0.009	0	0.006	0.009	0.009	0.009	0.009	0.01	0.009	0.009	0.01						
	L14	0		0.004	0.006	0.009	0.006	0	0	0	0	0.001	0	0	0.004	0.009					
	L15	0		0.025	0.022	0.026	0.022	0.022	0.022	0.022	0.022	0.024	0.022	0.022	0.022	0.026	0.022				
	L16	0		0.031	0.027	0.03	0.027	0.027	0.027	0.027	0.027	0.029	0.027	0.027	0.027	0.03	0.027	0.012			
	L17	0		0.012	0.007	0.004	0.007	0.01	0.01	0.01	0.01	0.009	0.01	0.01	0.012	0.004	0.01	0.027	0.032		
	L18	0		0.009	0.009	0	0.006	0.009	0.009	0.009	0.009	0.01	0.009	0.009	0.01	0	0.009	0.026	0.03	0.004	
	L19	0–0.001		0.004	0.007	0.008	0.007	0.001	0.001	0.001	0.001	0.002	0.001	0.001	0.005	0.008	0.001	0.023	0.028	0.011	0.008
EF1-α																					
	L1	0–0.011																			
	L2	0		0.012																	
	L3	0		0.035	0.024																
	L4	0		0.032	0.02	0.035															
	L5	0–0.011		0.024	0.017	0.042	0.027														
	L6	0		0.017	0.02	0.025	0.035	0.026													
	L7	0		0.021	0.024	0.04	0.035	0.032	0.01												
	L8	0		0.019	0.022	0.027	0.042	0.029	0	0.01											
	L9	0		0.024	0.026	0.026	0.036	0.034	0.017	0.021	0.019										
	L10	0		0.022	0.024	0.04	0.045	0.032	0.01	0.01	0.01	0.021									
	L11	0		0.015	0.019	0.045	0.039	0.024	0.02	0.024	0.022	0.026	0.024								
	L12	0		0.011	0.018	0.033	0.038	0.025	0.018	0.023	0.02	0.025	0.023	0.016							
	L13	0		0.03	0.02	0.015	0.03	0.037	0.03	0.035	0.032	0.027	0.034	0.04	0.028						
	L14	0		0.008	0.011	0.036	0.031	0.018	0.011	0.016	0.013	0.018	0.016	0.008	0.009	0.031					

Gene	Within lineages		Between lineages																	
	range	mean																		
L15	0		0.039	0.035	0.045	0.04	0.035	0.03	0.05	0.037	0.046	0.05	0.024	0.041	0.05	0.033				
L16	0		0.034	0.029	0.045	0.04	0.027	0.025	0.045	0.032	0.046	0.045	0.029	0.036	0.05	0.028	0.015			
L17	0		0.035	0.024	0.015	0.04	0.042	0.025	0.039	0.027	0.037	0.039	0.045	0.033	0.02	0.036	0.045	0.045		
L18	0		0.03	0.019	0.015	0.03	0.037	0.03	0.034	0.032	0.026	0.034	0.039	0.028	0	0.031	0.05	0.05	0.02	
L19	0		0.007	0.01	0.034	0.029	0.017	0.01	0.015	0.012	0.017	0.015	0.01	0.008	0.03	0.001	0.034	0.029	0.034	0.029

"/" indicates no data were available.

Chapter IV

Late Miocene-Pliocene diversification and Pleistocene-Holocene colonization shape phylogeography of the European *Lepidocyrtus lanuginosus* species group (Collembola: Entomobryidae)

Bing Zhang, Ting-Wen Chen, Eduardo Mateos, Stefan Scheu, Ina Schaefer

Abstract

Climatic, geographical and geological history are the main factors that determine the biogeography of present day biota. A growing number of phylogeographic studies have attempted to explain the origin and diversification of European biota, in order to better understand the mode and speed of speciation and the establishment of species. The European *Lepidocyrtus lanuginosus* species group (Collembola: Entomobryidae) is a monophyletic clade comprising two widely distributed species, but genetic analysis uncovered distinct lineages / cryptic species inhabiting different regions in Europe. We studied the phylogeography of this species group using a molecular multi-locus approach. Fragments of two mitochondrial (COI and COII) and two nuclear genes (28S rDNA D1–2 region and elongation factor 1-α, EF1-α) were used to infer the phylogeographic origin of the different lineages. We identified 18 lineages that did not overlap in their distribution range in Central, Southern, and Southeastern Europe, suggesting high genetic structure and limited gene flow between these three regions. Divergence date estimates based on the mitochondrial dataset and a relaxed molecular clock revealed that major lineages diverged in the Late Miocene and Pliocene (17–2.59 million years ago, Mya), i.e. before Quaternary ice ages, indicating that distinct lineages survived in multiple refugia in each sampling region during the Quaternary glacial periods. The genetic structure of the 18 lineages of the group supported a model of sequential allopatric diversification within each sampling region. Further, three distinct lineages which diverged during the Pleistocene and Holocene were widely distributed across Central Europe; suggesting that glacial cycles in the Quaternary affected the spread of these lineages. Identical haplotypes of both of these lineages, occurring in localities hundreds of kilometers apart, suggest recent human-mediated dispersal across Central Europe. The European *L. lanuginosus* species group is globally distributed and the lineages inhabiting other continents need to be included to understand its evolutionary history, radiation and distribution in a comprehensive way. The species complex provides an ideal model system to investigate the effects of historical, geographic, and climate processes as well as recent human activities on the evolution and biogeographic pattern of soil dwelling invertebrates across continents.

Keywords: springtail; *Lepidocyrtus cyaneus*; Quaternary ice ages; ribosomal subunit 28S rDNA D1–2 domain; elongation factor 1-α; cytochrome c oxidase subunit I and II

4.1 Introduction

Climatic, geographical, and geological history determine the present day biogeographic patterns of organisms that show a latitudinal diversity gradient, with species richness of most taxa increasing

towards the equator (Mittelbach et al., 2007). The palaeoecological record provided two broad and overlapping hypotheses on the localities where taxa survived glacial periods: (i) small and localised refugia in Southern European mountain areas, and (ii) broader, less well-defined, cryptic refugia of scattered populations (Provan and Bennett, 2008). Species' responses to climate change depend largely on their adaptations and environmental tolerances, resulting in individual responses in space and time (Stewart et al., 2010; Taberlet et al., 1998). Species of temperate regions likely survived glacial periods by retreating to southern refugia and recolonized higher latitude regions during warming periods (Hewitt, 2000, 1999; Hewitt et al., 1996). In contrast, cold-adapted species responded to climatic changes in the opposite way to temperate species, by retreating to refugia with small population sizes during periods of warm climate and expanding from refugia with large population sizes during cold phases (Dalén et al., 2005; Stewart and Dalén, 2008). Cryptic refugia for cold-adapted species were found north of Alps, e.g., the refugium area around the Carpathians for mammals (Kotlík et al., 2006; Sommer and Nadachowski, 2006), the adder (*Vipera berus*) (Ursenbacher et al., 2006), a moor frog (Babik et al., 2004), and newts (Babik et al., 2005), as well as polar refugia for bison (Shapiro et al., 2004) and the arctic fox (Dalén et al., 2005). Additionally, barrier effects by high mountains (e.g., the Pyrenees, Alps and Balkans) prevented exchange of biota (and genes) among these distinct regions and subsequently triggered an independent evolution of species and intraspecific lineages, e.g. in plants (Normand et al., 2011), vertebrates (Gassert et al., 2013), and invertebrates (Varga and Schmitt, 2008). Phylogeographic analyses of taxa inhabiting both sides of these barriers provide insight into the combined effects of historical climate change and geographic barriers on the present day genetic structure of species.

European Collembola comprise more than 2,500 species with species richness decreasing at higher latitude (Deharveng et al., 2008; Fiera and Ulrich, 2012; Ulrich and Fiera, 2009). Postglacial colonization from several glacial refugia was suggested as the main process that shaped the spatial distribution patterns of European and holarctic Collembola (Ávila-Jiménez and Coulson, 2011; Fiera et al., 2017; Fiera and Ulrich, 2012). However, many widely distributed European Collembola species show high intra-specific genetic variance (up to 20% in COI or COII), suggesting that they comprise a number of cryptic species, i.e., genetically distinct but morphologically undifferentiated lineages (Cicconardi et al., 2010; Porco et al., 2012a,b, 2014; Timmermans et al., 2005; von Saltzwedel et al., 2017, 2016; B. Zhang et al., 2018). According to divergence estimates of three Collembola species based on a strict molecular clock using mtDNA (Hebert et al., 2003; Papadopoulou et al., 2010), cryptic species or distinct lineages inhabiting different regions of Europe diverged in the Miocene (23.8-5.3 Mya) (von Saltzwedel et al., 2016), i.e., long before Quaternary glaciation. This suggests that Collembola do not follow the common pattern that species colonize from Southern (SE and SWE) to

Central Europe (CE) during Quaternary interglacial periods (Hewitt, 2000, 1999; Hewitt et al., 1996). Collembola colonize extreme habitats like deserts, high mountain soils and even the Antarctica (Hopkin, 1997), indicating that climate has a minor role on species richness and distribution compared to other arthropods, and lineages of Collembola survived Quaternary glaciation in refugia in Central Europe.

The European Collembola *Lepidocyrtus lanuginosus* species group comprises three morphospecies (Mateos, 2012), including *L. cyaneus* (dark blue body color) and *L. lanuginosus* (yellow body color) inhabiting both cold and warm regions across Europe (Cicconardi et al., 2010; Leinaas, 1981; Mateos, 2012; Salmon et al., 2014; https://fauna-eu.org/). However, DNA-based analyses revealed a gradient of intra-specific genetic distances of mitochondrial COI and COII (0-0.28) and discovered distinct lineages / cryptic species of both morphospecies inhabiting different regions of Europe (Cicconardi et al., 2010; Mateos et al., 2018; B. Zhang et al., 2018), indicating that this species group has evolved both temperate and cold-adapted lineages during the past millions of years. Both *L. cyaneus* and *L. lanuginosus* are paraphyletic but the species group is monophyletic (Mateos et al., 2018; B. Zhang et al., 2018), thus in all lineages of this species group were pooled to explore whether the Quaternary glacial cycles affect distribution and divergences of this complex. We investigated two hypotheses: (1) Lineages of the *L. lanuginosus* species group within sampling regions are phylogenetically clustered and isolated from lineages from other sampling regions, reflecting that lineages inhabiting CE originated from SE or SWE. (2) Divergent times of lineages of the *L. lanuginosus* species group fall into the Quaternary suggesting that Quaternary glaciation drove the dispersal and divergence of these lineages. For testing these hypotheses, we analyzed the genetic structure of these two morphospecies of the *L. lanuginosus* species group using cytochrome oxidase II (COII). The mitochondrial COI and two nuclear 28S rDNA D1−2 and EF1-α were also included and concatenated with COII to reconstruct a phylogeny and calculate the divergence times of all the lineages.

4.2 Material and methods

4.2.1 Data collection and lineage assignment

Nucleotide sequences of four genetic markers of the two morphospecies *L. cyaneus* and *L. lanuginosus* (Chapter II and IV; accession numbers available at NCBI GenBank; supplementary Table S4.1) were used to construct a phylogenetic tree (see Chapter III for details) with *Lepidocyrtus paradoxus* Uzel, 1890 and *L.* cf. *violaceus* Lubbock, 1873 as outgroups. For species collection and determination and assignment of genetic lineages, please refer to Material and Methods and Results in Chapter III. Lineage 2 of *L. cyaneus* was excluded from further analysis due to uncertainty in its

relationship to the other lineages as demonstrated by bootstrap values and Bayesian posterior probabilities (see Chapter III). The final tree included 224 individuals representing 18 lineages of the *L. lanuginosus* species group sampled across Europe. To analyze spatial genetic variance and structure, we assigned sampling localities to three geographical regions: Southern Europe (SE, south of the Alps, Italy and Southeastern France), South-western (SWE, Spain) and Central Europe (CE, north of the Alps) (Table 4.1).

4.2.2 Genetic diversity, population structure and haplotype reconstruction

We calculated standard genetic diversity indices for COII, assessed population structure within lineages through F_{ST} and performed AMOVA using sampling regions as geographical groups and lineages in DnaSP v.5.10 (Librado and Rozas, 2009) with 20,000 permutations. Further, we calculated the number of polymorphic sites (s), haplotype number (H), haplotype diversity (Hd) and nucleotide diversity (π) of each lineage and sampling region in DnaSP. The geographical distribution, genetic relationship, and the numbers of mutational steps among haplotypes of COII (alignment of 681 nucleotide characters and 152 accessions) were inferred by a TCS network using PopART v.1.7 (Leigh and Bryant, 2015).

4.2.3 Phylogenetic analyses and divergence time estimation

The four genetic markers were concatenated using SequenceMartrix v1.7.8 (Vaidya et al., 2011), with a final length of 2,242 bp including COI (651 bp), COII (681 bp), EF1-α (210 bp) and 28S rDNA (700 bp) sequences. Substitution models that best fitted each gene were defined according to the Bayesian Information Criterion (BIC) on IQ-TREE 1.6.6 (Nguyen et al., 2015) and were GTR+I+F+G (COI), HKY+I+F+G (COII), HKY+I+F (28S rDNA), and SYM+I (EF1-α). We estimated divergence times estimation using the concatenated alignment of all four loci in BEAST v.1.8.4 (Heled and Drummond, 2010) on online CIPRES services (Miller et al., 2010). Input files for BEAST v.1.8.4 were created with BEAUTi v.1.8.4 (Drummond et al., 2012). Runs were performed using the best fitted models under the assumption of an uncorrelated lognormal relaxed molecular clock using the estimated rate of 0.0169 ± 0.0019 substitutions/site/million years, according to the improved estimates of the mtDNA clock in insects (Papadopoulou et al., 2010). The number of generations for the total analysis was set to 600 million, with each chain sampled every 60,000 generations. The burn-in value was 10% and remaining

Table 4.1 Locality data of the lineages of the *Lepidocyrtus lanuginosus* species group and numbers of individuals of each population that sequenced for each gene. Loc. Abb.: Locality abbreviations.

Sampling region	Lineage	Locality	Loc. Abb.	Country	Numbers	28S D1-2	COII	COI	EF1-a
Central Europe (CE)									
	L3	Amiens	Ami	France	5	5	5	3	4
	L4	Antwerpen	Ant	Belgium	5	5	5	4	5
	L3	Bremen North	Bre	Germany	5	3	5	5	5
	L3	Chateau-Thierry	Cth	France	5	5	5	3	5
	L4	Illertissen	Ill	Germany	6	6	6	6	3
	L3/L13	Jena	Jen	Germany	9	9	8	7	8
	L3	Leipzig	Lei	Germany	5	5	5	4	4
	L3/L13	Ossenfeld arable land	Ossa	Germany	2	2	2	1	1
	L3	Ossenfeld grassland	Ossg	Germany	2	2	2	2	2
	L3	Oye-Plage (Calais)	Oyp	France	5	5	5	3	0
	L3	Schwechat	Sch	Austria	5	5	5	5	5
	L3	Sainte-Menehoult	Smen	France	5	5	5	5	5
	L4/L13	Utrecht	Utr	The Netherlands	8	7	8	5	8
	L3/L13/L15	Waake arable land	Waaa	Germany	8	8	8	5	5
	L13	Altmoorhausen	Alt	Germany	5	5	4	5	5
	L13	Billingshausen arable land	Billa	Germany	3	3	3	1	1
	L13	Billingshausen forest	Billf	Germany	5	5	5	1	1
	L13	Billingshausen grassland	Billg	Germany	4	4	4	1	1
	L13	Brno	Brn	Czech Republic	5	2	5	2	1
	L13	Deppoldshausen arable land	Depa	Germany	5	5	5	1	1
	L16	Deppoldshausen forest	Depf	Germany	5	5	5	2	2
	L13	Ellershausen arable land	Ella	Germany	3	3	3	0	0
	L16	Ellershausen forest	Ellf	Germany	5	5	5	0	0

Sampling region	Lineage	Locality	Loc. Abb.	Country	Numbers	28S D1-2	COII	COI	EF1-a
	L13	Herberhausen arable land	Hera	Germany	4	4	4	2	2
	L16	Herberhausen forest	Herf	Germany	5	5	5	1	1
	L15	Herberhausen grassland	Herg	Germany	5	5	5	4	4
	L13	Horka	Hor	Czech Republic	5	5	5	3	4
	L13	Kutno	Kut	Poland	5	4	5	5	5
	L13	Lebeny west	Lebe	Hungary	4	2	4	4	3
	L13	Nitra East	Niea	Slovakia	5	5	5	5	5
	L16	Ossenfeld forest	Ossf	Germany	3	3	3	1	1
	L13	Otterbach	Ott	Germany	5	5	5	4	5
	L13	Radom	Rad	Poland	5	5	5	5	5
	L18	Vichy	Vic	France	5	5	5	4	3
	L16	Waake forest	Waaf	Germany	5	5	5	2	2
Southern Europe (SE)									
	L5	Aiguafreda	Cas	Italy	5	5	5	4	4
	L6/7/8	Marseille-Allauch	Mal	France	5	5	5	3	4
	L10	Rapallo	Rap	Italy	5	5	5	4	2
	L11	Vernante	Ver	Italy	5	4	5	4	3
	L17	Park Gardasee	Pga	Italy	4	4	4	1	3
Southwestern Europe (SWE)									
	L1	Tagamanent	Tag	Spain	5	5	5	2	3
	L9	Pineta valley	Pin	Spain	5	5	5	2	5
	L1/12	Aiguafreda	Aig	Spain	4	4	4	4	4
	L12/14	Serra de Catllaràs	Cat	Spain	5	5	5	5	5
	L1	Montnegre	Mon	Spain	5	5	5	5	4
	L19	Viladecavalls	Vil	Spain	5	5	5	2	4
	Total				224	214	222	147	153

parameters were set as default. Chains were run in three independent times and subsequently combined using LogCombiner v.1.8.4 (http://beast.community/index.html). To confirm convergence, the effective sample size (ESS) values were checked in Tracer 1.6 (Rambaut et al., 2013). The trees were sampled at every 60,000 steps and then summarized in a maximum clade credibility tree using TreeAnotator v.1.8.4 (Drummond et al., 2012). The final tree was visualized in FigTree v.1.4.2 (Rambaut, 2012).

4.3 Results

4.3.1 Genetic diversity

In total, we used 222 COII sequences from 224 specimens of 18 lineages to study the genetic diversity of the *L. lanuginosus* species group (Table 4.1). We recorded a large number of haplotypes (93) and high haplotype (0.97 ± 0.01) and nucleotide diversity (0.17 ± 0.004, Table 4.2). The majority of the sequences (169, 76%) and haplotypes (55, 59%) were collected in CE. The highest haplotype and nucleotide diversity was found in SE (0.98 and 0.15, respectively), followed by SWE (0.95 and 0.14, respectively), and CE (0.95 and 0.13, respectively) (Table 4.2).

About 63% of the sequences were assigned to L3, L13 and L16, all occurring in CE and comprising 46, 71, and 23 sequences, and 16, 13, and 11 haplotypes, respectively (Table 4.1 and 4.2). Of these three lineages, L13 had the highest nucleotide diversity (0.03) but lowest haplotype diversity (0.78). Lineage L3, conversely, had the lowest nucleotide diversity (0.01) but the highest haplotype diversity (0.91).

4.3.2 Genetic structure

No lineages were shared among the three sampling regions (Figure 4.1). Six lineages (L3–4, L13, L15–16 and L18) were collected from CE, seven (L5–8, L10–11 and L17) from SE, and five (L1, L9, L12, L14 and L19) from SWE. Distinct lineages coexisted in populations within each sampling region (Table 4.1), for example, L3 and L13 co-occurred in Germany and the Netherlands, CE; L6–8 in France, SE; and L1 and L12 in Spain, SWE. In CE all lineages except L18, were found in at least three different localities. Lineages in SE were all local and endemic. In SWE three lineages (L9, L14 and L19) were endemic but L1 and L12 were found in at least two localities.

For all 18 lineages, genetic variability within lineages was extremely low (1.6%) as compared to the variability among lineages (98.4%) (Table 4.3), and all lineages were clearly separated from each

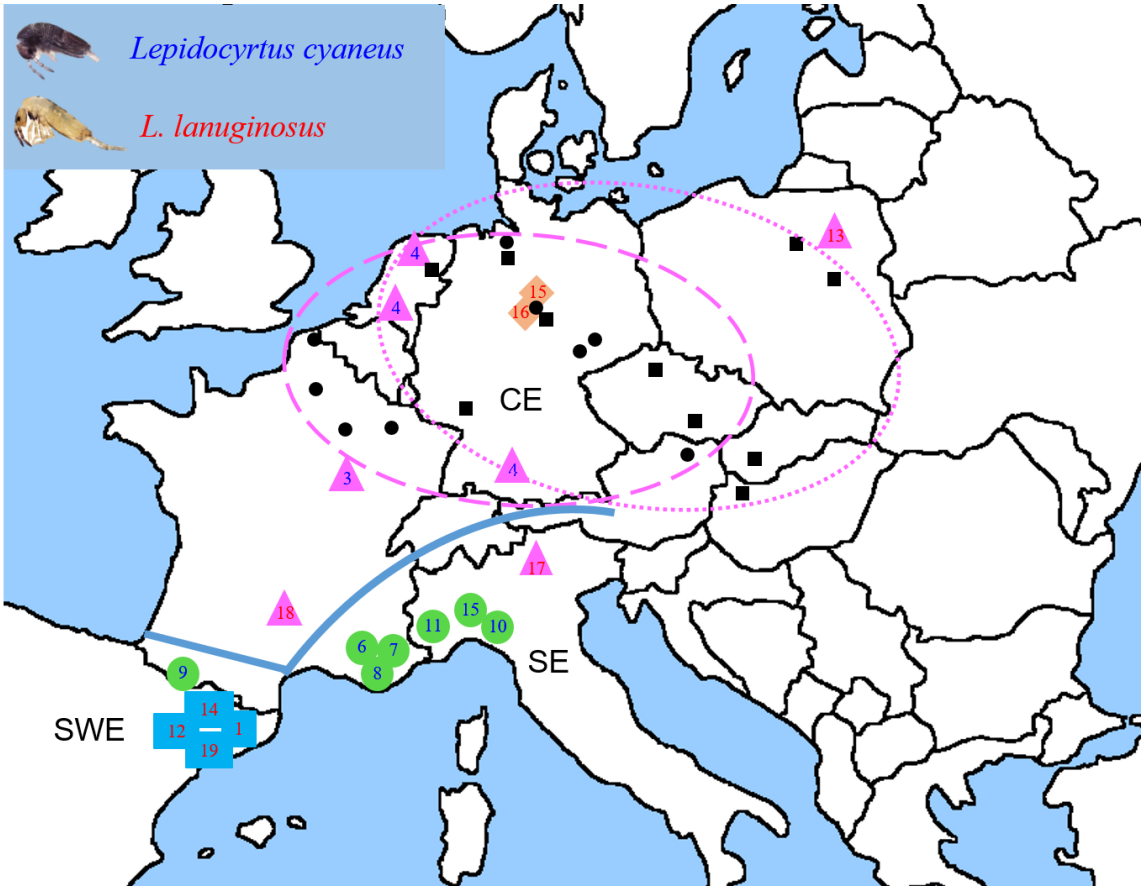


Figure 4.1 Map with sampling locations of all 18 lineages of the *Lepidocyrtus lanuginosus* species group. The light blue thick lines separate sampling regions: CE, Central Europe; SE, Southern Europe; and SWE, Southwestern Europe. Numbers in read indicate lineages of *L. lanuginosus*, numbers in blue those of *L. cyaneus*. The 18 lineages were assigned to four clades marked with different colors and shapes: Clade I-IV which refers to Figure 4.3, yellow diamonds, green rounds, blue squares, and pink triangles, respectively.

other (TSC network, Figure 4.2). The geographic structure of the 46 populations was weak, with the genetic variability (~60%) within regions among populations being almost twice as high as that among regions (~31%) (Table 4.3). However, the geographic structure within each region was strong, with genetic variability among populations of ~86%, ~89%, and ~82% in CE, SE, and SWE, respectively (Table 4.3). Fu's Fs was significant for all three regions but Tajima's D was not significantly different from 0, suggesting selective neutrality of the observed nucleotide polymorphisms.

Lineages L3, L4, and L13 were widely distributed in CE, in particular L3 and L13 that ranged from the west of France and the Netherlands to Germany and Austria (Figure 4.1). The geographic structure of L3 was weak; most of the genetic variability was within populations (86.7%) rather than among populations (13.3%) (Table 4.3), with seven out of the 11 populations comprising more than three haplotypes in the five individuals studied (supplementary Table S4.2 & Figure S4.1a); haplotype Lcya_Ami_1 was distributed the widest, occurring in Amiens, Oye-Plage and Sainte-Menehoult, France,

Table 4.2 Genetic diversity indices for lineages of the European *Lepidocyrtus lanuginosus* species group, calculated according to genetic lineages and geographic regions.

Lineage/region	N	ss	Eta	h	Hd (SD)	π (SD)	k	D*	F*	*Fs*	D	F_{ST}
L1	12	9	9	7	0.879 (0.075)	0.00454 (0.00059)	3.091	-0.32484	-0.22958	-1.281	0.15071	
L3	46	31	33	16	0.913 (0.019)	0.01289 (0.00049)	8.776	-0.07795	0.17965	0.523	0.57301	0.13269
L4	16	23	23	6	0.800 (0.068)	0.01189 (0.00124)	8.1	-0.13007	0.1139	3.85*	0.68552	
L5	5	27	27	5	1.000 (0.126)	0.02026 (0.00438)	13.8	0.48338	0.52024	0.075	0.48338	
L6	1			1								
L7	2	10	10	2	1.000 (0.500)	0.01468 (0.00734)	10					
L8	2	1	1	2	1.000 (0.500)	0.00147 (0.00073)						
L9	5	12	13	4	0.900 (0.161)	0.01057 (0.00269)	7.2	1.35005	1.38429	1.072	1.11728	
L10	5	27	28	5	1.000 (0.126)	0.01689 (0.00810)	11.5	-1.02194	-1.11235	-0.156	-1.07748	
L11	5	32	32	3	0.800 (0.164)	0.02261 (0.00675)	15.4	0.0195	0.02103	4.846	0.0195	
L12	3	11	11	3	1.000 (0.272)	0.01077 (0.00295)	1					
L13	71	78	83	13	0.779 (0.034)	0.02951 (0.00160)	20.096	0.94721	0.95924	15.559**	0.57631	0.78532***
L14	4	5	5	3	0.833 (0.222)	0.00441 (0.00130)	3	0.95621	0.90358	0.731	0.95621	
L15	8	7	7	5	0.786 (0.151)	0.00257 (0.00085)	1.75	-1.82716*	-1.97974	-1.449	-1.67405	
L16	23	12	12	11	0.877 (0.049)	0.00293 (0.00039)	1.992	-1.11988	-1.38072	-5.381**	-1.34232	
L17	4	2	2	2	0.500 (0.265)	0.00147 (0.00078)	1	-0.7099	-0.60427	1.099	-0.7099	
L18	5	10	10	4	0.900 (0.161)	0.00793 (0.00209)	5.4	0.8945	0.94334	0.612	0.8945	
L19	5	0	0	1								
CE	169	261	371	55	0.951 (0.008)	0.14227 (0.00471)	96.884	2.49809**	2.45589**	29.623**	1.58491	0.86216***
SE	24	235	327	20	0.982 (0.018)	0.15489 (0.00528)	105.478	1.45217*	1.47268	2.614	0.82488	0.88849***
SWE	29	217	274	18	0.951 (0.024)	0.13216 (0.00856)	90	1.79438**	1.86109**	9.468**	1.13192	0.81920***
Total	222	300	488	93	0.970 (0.005)	0.16510 (0.00353)	112.432	2.61078**	2.25261**	12.519**	1.20018	0.88629***

N, number of sequences; h, number of haplotypes; ss, number of segregating sites; Eta, total number of mutations; Hd, haplotype diversity; π, nucleotide diversity; k, average number of pairwise nucleotide differences; D*, Fu and Li's D* test statistic; F*, Fu and Li's F* test statistic; Fs, Fu's Fs statistic; D, Tajima's D; SD, standard deviation.*0 .02 < p < 0.05; **0.001 < p < 0 .02, ***P < 0.01. CE = Central Europe; SE = Southern Europe; SWE = Southwestern Europe.

Table 4.3 Results of the AMOVA on the variation in COII among sampling regions and lineages, among populations, within populations and lineages and in total. Populations with less than two individuals were excluded from the analysis.

Sum of Variation	df	squares	sigma^2	% variation
All 46 populations				
Among regions	2	306076.397	2878.541	31.12914
Among populations	43	1181513.739	5530.662	59.80973
Within populations	176	147468.842	837.891	9.06113
All	221	1635058.977	9247.094	
All populations in CE				
Among populations	34	921711.773	5283.964	86.2164
Within populations	134	113197.642	844.759	13.7836
All	168	1034909.414	6128.723	
All populations in SE				
Among populations	4	133750.567	5458.76	88.84926
Within populations	19	13016.600	685.084	11.15074
All	23	146767.167	6143.844	
All populations in SWE				
Among populations	5	126051.400	4187.270	81.92049
Within populations	23	21254.600	924.113	18.07951
All	28	147306.000	5111.383	
All 18 lineages				
Among lineages	17	1610535.839	7245.461	98.36795
Within lineages	204	24523.138	120.211	1.63205
All	221	1635058.977	7365.672	
All populations of L3				
Among populations	10	742.639	6.669	13.26887
Within populations	35	1525.600	43.589	86.73113
All	45	2268.239	50.257	
All populations of L13				
Among populations	17	16261.241	214.961	78.53152
Within populations	53	3114.533	58.765	21.46848
All	70	19375.775	273.726	

CE = Central Europe; SE = Southern Europe; SWE = Southwestern Europe.

and Bremen, Jena, and Leipzig, Germany (supplementary Figure S4.1a). By contrast, the geographic structure of L13 was very strong with most of the genetic variability among populations (78.5%, Table 4.3) and the majority of the populations comprising only one haplotype (supplementary S4.2 & Figure S4.1b); haplotype Llan_Hera_1 was distributed the widest, occurring in Utrecht, The Netherlands,

Waake and Herberhausen, Germany, Kutno and Radom, Poland, and Nitra, Slovakia (supplementary Figure S4.1b).

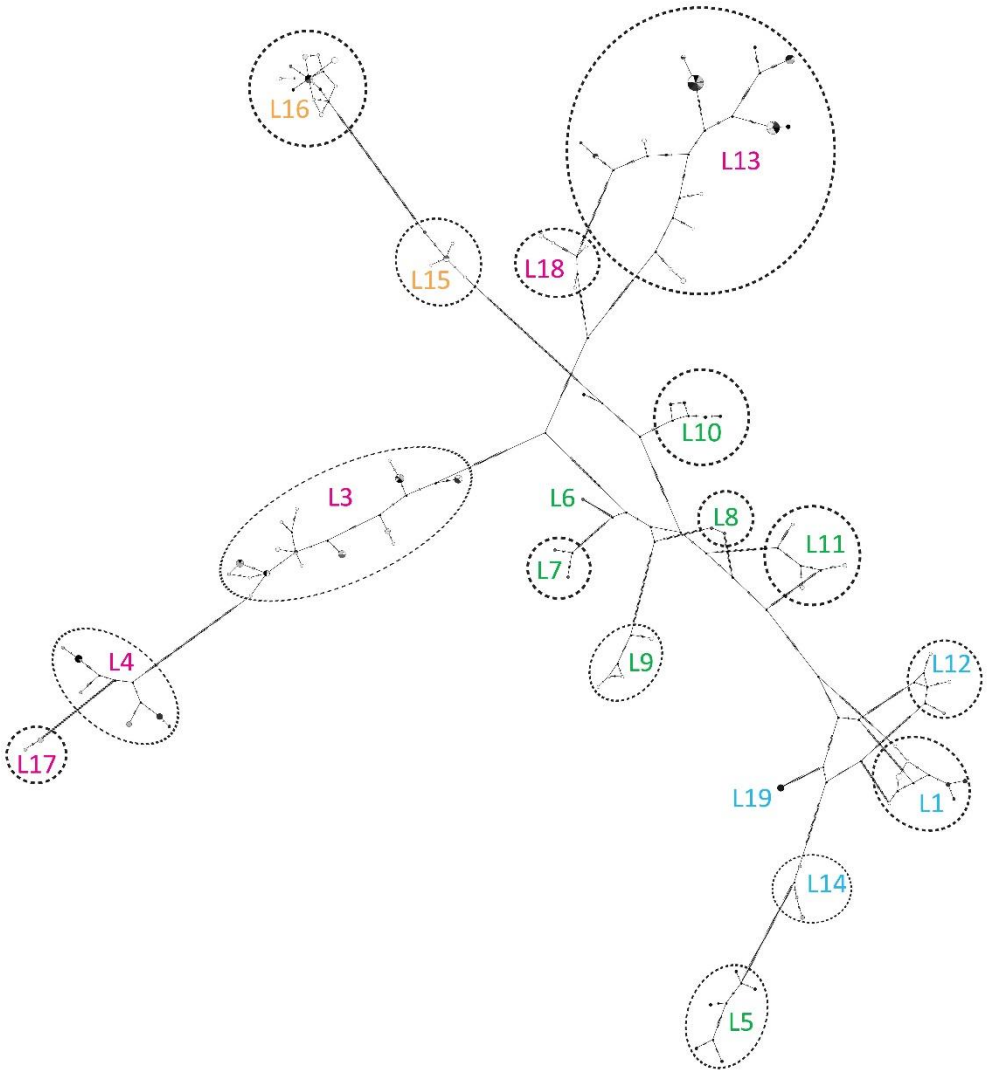


Figure 4.2 TCS haplotype network for COII of all 18 lineages of the *Lepidocyrtus lanuginosus* species group. Solid circle size is proportional to the frequency of each lineage and dashes represent mutational events. The 18 lineages were assigned to four clades marked with different colors: Clade I-IV refer to Figure 4.3, yellow, green, blue, and pink, respectively.

4.3.3 Phylogenetic inferences and divergence time estimation

The 18 lineages were grouped into four monophyletic clades; major divergences between and within clades occurred during the Late Miocene (5.3–11.3 Mya, Figure 4.3). The oldest split in the *L. lanuginosus* species group separated Clade I (L15 and L16 in CE) from all remaining clades (Clade II–IV) and was dated to the Middle Miocene (8.62–16.89 Mya).

All 13 lineages of Clade II and III were from SE and SWE, and separated from Clade IV during the Late Miocene (8.07–15.57 Mya). Clade II and III also separated during the Late Miocene (6.85–13.56

Mya), and lineages in these two clades diverged from the Late Miocene to the Pliocene (1.59–11.79 Mya).

All five lineages of Clade IV were from CE, except L17 which was from SE (Figures 4.1 and 4.3) and separated from the CE clades during the late Miocene. The oldest split was between L4 and the remaining lineages (6.33–13.19 Mya), followed by the split of L17 and L3, L13, and L18 (5.27–11.57 Mya). The most diverse and widely distributed lineages L3 and L13 separated between 3.96 and 8.48 Mya.

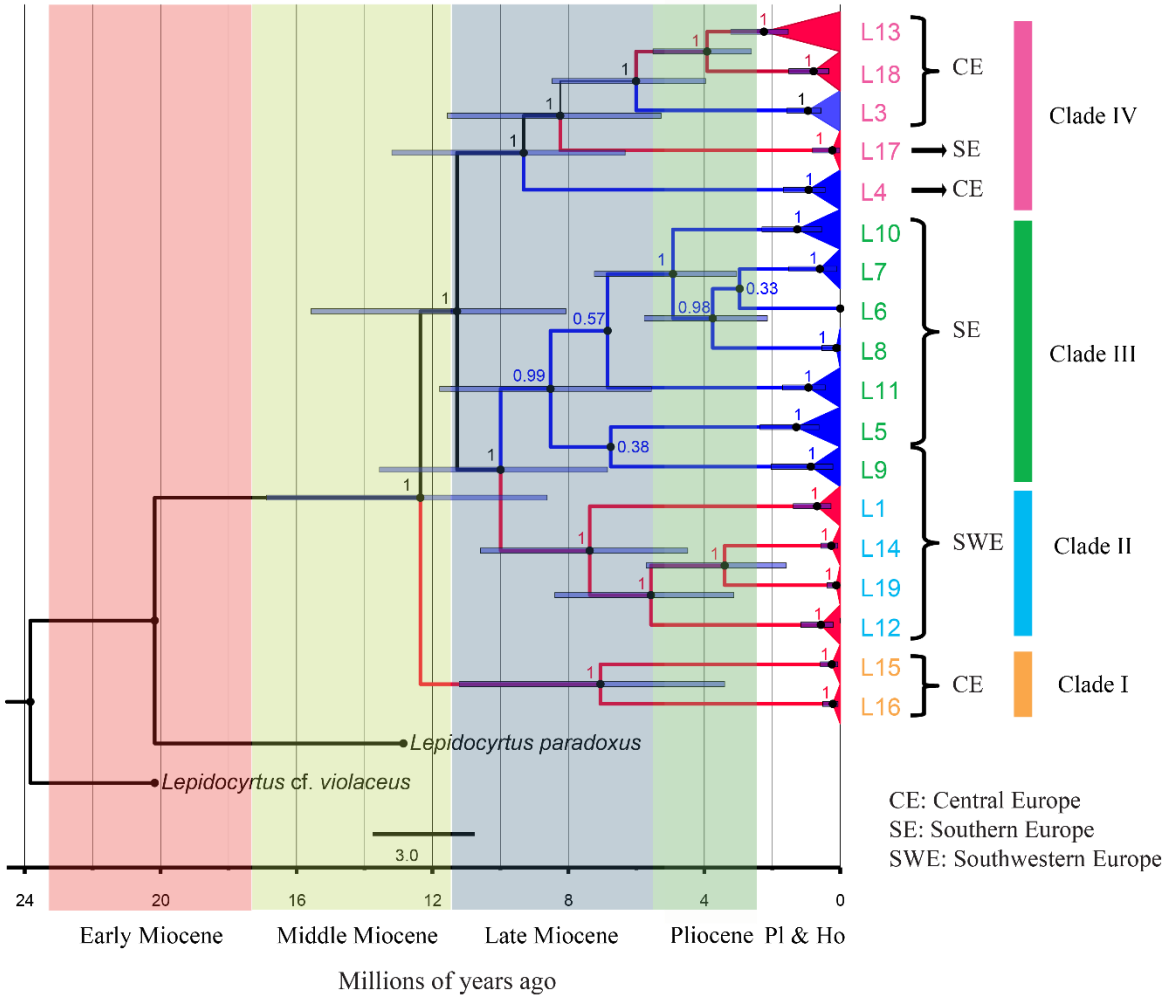


Figure 4.3 Phylogenetic relationships and molecular divergence time estimates of 18 lineages (L1, L3-19) of the *Lepidocyrtus lanuginosus* species group in Europe. The Bayesian phylogenetic tree is based on nucleotide sequence data of four genes (28Sr DNA, EF1-α, COI and COII). Molecular divergence time estimates based on COI (651 bp) and COII (681 bp) sequences calculated with BEAST. Blue bars on nodes represent divergence times with 95% confidence intervals in millions of years, Mya. A geological timescale is provided below the tree, indicating the Miocene (23.03–5.33 Mya), the Pliocene (5.33–2.59 Mya), the Pleistocene (Pl, 2.59–0.0117 Mya), and the Holocene (Ho, 0.0117 to present). Terminal clades are collapsed, following by lineage names. The red branches indicate lineages of *L. lanuginosus*, the blue ones lineages of *L. cyaneus*. Numbers on nodes are posterior probabilities. The 18 lineages were assigned to four clades marked with different colors: Clade I-IV, yellow, green, blue, and pink, respectively.

4.4 Discussion

4.4.1 Biogeographic distribution pattern

We detected 18 lineages belonged to two morphospecies, *L. cyaneus* and *L. lanuginosus* (Chapter III), which were previously thought to be two of the most widespread and abundant species of Collembola in Europe (https://fauna-eu.org/). But genetic data provided an incongruent perspective on the biogeography of these two morphospecies, indicating that distinct lineages of each morphospecies have very limited range size, especially in SE and SWE (Figure 4.1). The European *L. lanuginosus* species group comprises four monophyletic clusters with two of them colonizing more than one geographic region. The pattern suggests that lineages of the *L. lanuginosus* species group in at least three of the regions are non-monophyletic and include both morphospecies *L. cyaneus* and *L. lanuginosus* which are non-monophyletic (Figure 4.2). Thus we reject our first hypothesis, but rather suggest the existence of multiple colonization in each of the three regions (SE, SWE, and CE) by lineages of the *L. lanuginosus* species group. This study suggests that the molecular phylogeny of widely distributed European Collembola species differs from phylogenies based on morphological characters (Fiera et al., 2017). Therefore, the integration of genetic and morphological data is important for more accurate understanding of the spatial distribution and genetic structure of Collembola.

Lineages inhabiting CE were never found in SE and SWE, suggesting that the Pyrenees and Alps function as dispersal barriers preventing dispersal of Collembola from CE to SE or SWE and vice versa. Notably, this pattern opposes the common perspective that European species retreated to southern refugia during Quaternary glaciation and recolonized CE thereafter (Hewitt, 2000, 1999; Hewitt et al., 1996). Distinct lineages of the Collembola species *Orchesella cincta* which occurs in different regions of Europe also suggest that the Alps form a strong barrier to gene flow between Italy and CE (Timmermans et al., 2005). Collembola evolved a number of strategies to adapt to low temperature, including freeze avoidance, freeze tolerance and desiccation resistance (Ávila-Jiménez *et al.* 2010; Teets and Denlinger 2014). Many Collembola species, including *L. cyaneus*, are found under snow, e.g., in Norwegian spruce forests (Hågvar, 2010; Leinaas, 1981) and Northeastern China (Zhang et al., 2014). Additionally, during glacial periods, Collembola, including *L. lanuginosus*, may have avoided low temperature conditions and frost by moving deeper into the soil (Gass et al., 2006; Healey, 1967). Thus, regions north of the Alps likely allowed the six lineages detected in CE to survive Quaternary glaciation north of the Alps, along with other Collembola species (Timmermans et al., 2005; von Saltzwedel et al., 2017, 2016). However, the six lineages in CE were non-monophyletic (Figure 4.2), indicating that multiple lineages colonized CE with this colonization presumably occurring in the

middle Miocene and shaping the present day genetic structure of the European *L. lanuginosus* species group. This agrees that multiple colonization from different refugia have shaped the spatial distribution of European Collembola, butterflies and beetles (Fattorini and Ulrich, 2012; Fiera et al., 2017; Fiera and Ulrich, 2012; Habel et al., 2011). A larger sampling area across Northern Europe and Russia are needed to substantiate the origins of the six lineages and their phylogenetic relationships.

The number of lineages / cryptic species in SE and SWE (12) were twice as high in CE (6) and this complies with the latitudinal gradient of increasing Collembola diversity towards the Balkans and the Mediterranean region (Ulrich and Fiera, 2009). Similar to other Collembola species (von Saltzwedel et al., 2016), both SE and SWE were rich in endemic lineages of the *L. lanuginosus* species group, suggesting that the true lineage richness in these regions is likely to be underestimated. DNA-based taxonomy has identified gaps in genetic distances (Sun et al., 2017; F. Zhang et al., 2018) and suggests that the lineages of the European *L. lanuginosus* species group resemble different species (Chapter III). This further supports the notion that the true diversity of Collembola is largely underestimated (Cicconardi et al., 2010; Deharveng, 2004). Indeed, the Mediterranean region is rich in endemic Collembola species with many still undescribed (Deharveng et al., 2008; Fiera and Ulrich, 2012).

4.4.2 Late Miocene-Pliocene diversification across Europe

Results of the present study suggest that the four monophyletic clades of the European *L. lanuginosus* species group diverged during the Miocene (6.85–16.89 Mya, Figure 4.2), thus rejecting our second hypothesis. Global climate has undergone significant and complex changes throughout the Cenozoic (65.5–0 Mya), with gradual trends of cooling and warming, and glacial cycles during the Miocene, Pliocene and Pleistocene (Louwye et al., 2008; Zachos et al., 2001). The four clades of the European *L. lanuginosus* species group separated in the late Miocene when the global climate changed from warm and humid to cool and dry (Zachos et al., 2001). The general cooling trend in the late Miocene caused increased seasonality in Europe (15.97–7.25 Mya), while the simultaneous decrease of atmospheric CO_2 concentration and increase of fire frequency contributed to the expansion of C_4 grasslands in Europe (Beerling and Osborne, 2006; Bruch et al., 2007; Mosbrugger et al., 2005). Most lineages in the present study were sampled from grasslands (supplementary Table S4.1), so their common ancestor may have colonized the three sampling regions with the expansion of grasslands.

Additionally, in contrast to the strict isolation by distance in *O. cincta* (van der Wurff et al., 2005), genetically similar lineages are not always geographically close, for example, L17 is genetically closer to lineages in CE but geographically closer to lineages in SE, and L9 is genetically closer to lineages in SE but geographically closer to lineages in SWE (Figures 4.1 and 4.3). Clade IV diverged from Clade II

and III and colonizes both SE (L17) and CE (all other lineages). This pattern suggests colonization of Italy from CE by Clade IV, i.e. from north to south, rather than commonly assumed from south to north (Figure 4.2). These complicated colonization events may relate to the past climate changes as environmental factors were indicated to play a significant role in shaping the species or even lineages distribution of Collembola (Kováč et al., 2016; Salmon et al., 2014; B. Zhang et al., 2018).

Lineages inhabiting SE and SWE comprised two monophyletic clades (Clade II and III), which also diverged in the late Miocene, as indicated by our molecular clock dating. The Mediterranean region is rich in cryptic species of *L. lanuginosus*, comprising numerous distantly related genetic lineages colonizing sites in close vicinity to each other (Cicconardi et al., 2010). The divergence times of L6-8 and L10 in SWE and L12, L14, and L19 in SE dated to the Messinian salinity crisis (MSC, between 5.59 and 5.33 Mya) of the Mediterranean region. At this time, the sea level of the Mediterranean Sea dropped by about 1,000 m below the current level, transforming the basin into a continuous landmass (Clauzon et al., 1996; Krijgsman et al., 1999), which may have facilitated the spread of Collembola across the Mediterranean region and contributed to later split up into different lineages. However, Cicconardi et al. (2010) may have underestimated the divergence time of Mediterranean lineages of the *L. lanuginosus* species group by using a mutation rate for COII of 4.96% per Ma, more than twice that used by von Saltzwedel et al. (2016) and in this study (~ 2.3% per Ma). Nevertheless, our results support the conclusion that sea level changes during the MSC shaped the distribution of genetic lineages of morphospecies of *Lepidocyrtus* and contributed to their high intra-specific genetic divergence (Cicconardi et al., 2010).

4.4.3 Pleistocene-Holocene colonization patterns in Central Europe

Three lineages, i.e., L3, L4, and L13, were widely distributed in CE with high haplotype diversity (Table 4.2 and Figure 4.1), similar to the three main lineages of *Parisotoma notabilis* which are distributed across Europe (Porco et al., 2012b; von Saltzwedel et al., 2017). The common ancestor of each main lineage was dated to the Pleistocene (Figure 4.2), suggesting that Quaternary glaciation likely drove the expansion and isolation of these lineages. Collembola are small and wingless hexapods with presumed limited dispersal ability (Bengtsson et al., 1994; Hopkin, 1997). However, some winter-active species utilize the snow surface for long distance dispersal (Hågvar, 2000, 1995; Zhang et al., 2017). The *L. lanuginosus* species group inhabits both cold and warm environments and the ancestor of the cold-adapted or cold-tolerant lineages likely was able to survive or even expand its range during Quaternary glaciation of CE. Some species of Collembola are able to survive frost conditions of below -20°C for more than four years (Coulson and Birkemoe, 2000) and in seawater for at least two weeks

(Coulson et al., 2002), and far-flung dispersal in ice packs through the Arctic has been suggested (Moore, 2002). Ocean current patterns have been suggested as important factor shaping the distribution of Arctic and Antarctic Collembola (Ávila-Jiménez and Coulson, 2011; McGaughran et al., 2010). Due to the remarkable ability of Collembola to survive and disperse at low temperature conditions species / lineages inhabiting CE today may have colonized CE from northern Europe or northern Asia during Quaternary glacial times.

Additional to ancient divergences and distributions, humans also likely shaped the present day distribution of Collembola in Europe. Identical and closely related haplotypes of the three widespread lineages L3, L4, and L13 originated from grasslands hundreds of kilometers apart (supplementary Figures S4.1a, b), suggesting that human activity associated with grassland management may have facilitated the dispersal of lineages of the *L. lanuginosus* species group across Europe. DNA barcode analysis has revealed multiple introduction of five Collembola species between Europe and North America (Porco et al., 2013). Further, Cicconardi et al. (2017) recently showed that almost 30% of the Collembola species from Mascarene Islands in the Indian Ocean are genetically identical or near identical to individuals from very distant geographic regions of the world, suggesting that they were spread by humans. Thus the present day distribution pattern of soil-living Collembola may be strongly affected by human trading and travelling. This likely contributed to high haplotype diversity of populations in and near to cities. In fact, in the present study haplotype diversity was at a maximum in sampling sites close to cities such as Bremen and Brno. More sites need to be sampled to reconstruct dispersal routes of Collembola in Europe and to uncover their origin, spread and radiation.

4.5 Conclusions

The structure of genetic lineages of the European *L. lanuginosus* species group suggests multiple colonization events of species / lineages across CE, SE and SWE. Divergence time estimation of the four clades identified date to the middle and late Miocene with each clade further diversifying during the Pliocene. The late Miocene divergence of clades may have been due to declining temperature and expansion of grasslands in Europe, while Pliocene divergences in SE and SWE may have been favored by sea level changes during the Messinian salinity crisis. SE and SWE are rich in endemic lineages suggesting that the Pyrenees and Alps functioned as dispersal barriers. By contrast, a number of lineages in CE are widely distributed potentially due to population expansion during Quaternary interglacial periods or recent human-mediated dispersal. Overall, the results indicate that colonization of Europe by soil-living invertebrates such as Collembola is more complex than previously assumed and markedly differs from that of plants and mammals. Species of the *L. lanuginosus* species group

are globally distributed, genetic comparison with these two morphospecies from outside Europe may provide additional evidence on how the past climate changes shape the dispersal and divergence of this species group. Thus we call for global scale phylogeographic studies on the present day globally distributed Collembola and other soil invertebrates to uncover their dispersal, speciation, and evolution.

Acknowledgements

We thank Andreas Klein and Bastian Heimburger for the help in sample collection. The study was supported by the German Ministry of Education, Science and Technology (BMBF). Bing Zhang was supported by the China Scholarship Council.

References

Ávila-Jiménez, M.L., Coulson, S.J., 2011. A holarctic biogeographical analysis of the Collembola (Arthropoda, Hexapoda) unravels recent post-glacial colonization patterns. Insects 2, 273–296. doi:10.3390/insects2030273

Ávila-Jiménez, M.L., Coulson, S.J., Solhøy, T., Sjöblom, A., 2010. Overwintering of terrestrial Arctic arthropods: the fauna of Svalbard now and in the future. Polar Res. 29, 127–137. doi:10.1111/j.1751-8369.2009.00141.x

Babik, W., Branicki, W., Crnobrnja-Isailović, J., Cogălniceanu, D., Sas, I., Olgun, K., Poyarkov, N.A., Garcia-París, M., Arntzen, J.W., 2005. Phylogeography of two European newt species - Discordance between mtDNA and morphology. Mol. Ecol. 14, 2475–2491. doi:10.1111/j.1365-294X.2005.02605.x

Babik, W., Branicki, W., Sandera, M., Litvinchuk, S., Borkin, L.J., Irwin, J.T., Rafiński, J., 2004. Mitochondrial phylogeography of the moor frog, *Rana arvalis*. Mol. Ecol. 13, 1469–1480. doi:10.1111/j.1365-294X.2004.02157.x

Beerling, D.J., Osborne, C.P., 2006. The origin of the savanna biome. Glob. Chang. Biol. 12, 2023–2031. doi:10.1111/j.1365-2486.2006.01239.x

Bengtsson, G., Hedlund, K., Rundgren, S., 1994. Food- and density-dependent dispersal: evidence from a soil collembolan. J. Anim. Ecol. 63, 513–520. doi:10.2307/5218

Bruch, A.A., Uhl, D., Mosbrugger, V., 2007. Miocene climate in Europe — Patterns and evolution. Palaeogeogr. Palaeoclimatol. Palaeoecol. 253, 1–7. doi:10.1016/j.palaeo.2007.03.030

Cicconardi, F., Borges, P.A. V., Strasberg, D., Oromí, P., López, H., Pérez-Delgado, A.J., Casquet, J., Caujapé-Castells, J., Fernández-Palacios, J.M., Thébaud, C., Emerson, B.C., 2017. MtDNA metagenomics reveals large-scale invasion of belowground arthropod communities by introduced species. Mol. Ecol. 26, 3104–3115. doi:10.1111/mec.14037

Cicconardi, F., Nardi, F., Emerson, B.C., Frati, F., Fanciulli, P.P., 2010. Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. Mol. Ecol. 19, 386–400. doi:10.1111/j.1365-294X.2009.04457.x

Clauzon, G., Suc, J.-P., Gautier, F., Berger, A., Loutre, M., 1996. Alternate interpretation of the Messinian salinity crisis: Controversy resolved? Geology 24, 363. doi:10.1130/0091-7613(1996)024<0363:AIOTMS>2.3.CO;2

Coulson, S.J., Birkemoe, T., 2000. Long-term cold tolerance in Arctic invertebrates : recovery after 4 years at below – 20 ° C. Can. J. Zool. 78, 2055–2058. doi:10.1139/cjz-78-11-2055

Coulson, S.J., Hodkinson, I.D., Webb, N.R., Harrison, J. a., 2002. Survival of terrestrial soil-dwelling arthropods on and in seawater: implications for trans-oceanic dispersal. Funct. Ecol. 16, 353–356. doi:10.1046/j.1365-2435.2002.00636.x

Dalén, L., Fuglei, E., Hersteinsson, P., Kapel, C.M.O., Roth, J.D., Samelius, G., Tannerfeldt, M., Angerbjörn, A., 2005. Population history and genetic structure of a circumpolar species: the arctic fox. Biol. J. Linn. Soc. 84, 79–89. doi:10.1111/j.1095-8312.2005.00415.x

Deharveng, L., 2004. Recent advances in Collembola systematics. Pedobiologia. 48, 415–433. doi:10.1016/j.pedobi.2004.08.001

Deharveng, L., D'Haese, C.A., Bedos, A., 2008. Global diversity of springtails (Collembola; Hexapoda) in freshwater. Hydrobiologia 595, 329–338. doi:10.1007/s10750-007-9116-z

Drummond, A.J., Suchard, M. a, Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 3–6. doi:10.1093/molbev/mss075

Fattorini, S., Ulrich, W., 2012. Spatial distributions of European Tenebrionidae point to multiple postglacial colonization trajectories. Biol. J. Linn. Soc. 105, 318–329. doi:10.1111/j.1095-8312.2011.01797.x

Fiera, C., Habel, J.C., Kunz, M., Ulrich, W., 2017. Large-scale phylogenetic structure of European springtails (Collembola) depends on species range size and postglacial colonization history. Biol. J. Linn. Soc. 120, 664–674. doi:10.1111/bij.12907

Fiera, C., Ulrich, W., 2012. Spatial patterns in the distribution of European springtails (Hexapoda: Collembola). Biol. J. Linn. Soc. 105, 498–506. doi:10.1111/j.1095-8312.2011.01816.x

Gass, F., Gillet, S., Ponge, J.-F., 2006. The use of directional traps for the assessment of short-term phenanthrene effects upon soil springtail communities. Environ. Pollut. 140, 364–370. doi:10.1016/j.envpol.2005.06.027

Gassert, F., Schulte, U., Husemann, M., Ulrich, W., Rödder, D., Hochkirch, A., Engel, E., Meyer, J., Habel, J.C., 2013. From southern refugia to the northern range margin: genetic population structure of the common wall lizard, *Podarcis muralis*. J. Biogeogr. 40, 1475–1489. doi:10.1111/jbi.12109

Habel, J.C., Lens, L., Rödder, D., Schmitt, T., 2011. From Africa to Europe and back: refugia and range shifts cause high genetic differentiation in the Marbled White butterfly *Melanargia galathea*. BMC Evol. Biol. 11, 215. doi:10.1186/1471-2148-11-215

Hågvar, S., 2010. A review of Fennoscandian arthropods living on and in snow. Eur. J. Entomol. doi:10.14411/eje.2010.037

Hågvar, S., 2000. Navigation and behaviour of four Collembola species migrating on the snow surface. Pedobiologia. 44, 221–233.

Hågvar, S., 1995. Long distance, directional migration on snow in a forest collembolan, *Hypogastrura socialis* (Uzel). Acta Zool. Fenn. 196, 200–205.

Healey, I.N., 1967. An ecological study of temperatures in a Welsh moorland soil. J. Anim. Ecol. 36, 425–434. doi:10.2307/2923

Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B Biol. Sci. 270, 313–321. doi:10.1098/rspb.2002.2218

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570–580. doi:10.1093/molbev/msp274

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. Nature 405, 907–913. doi:10.1038/35016000

Hewitt, G.M., 1999. Post-glacial re-colonization of European biota. Biol. J. Linn. Soc. 68, 87–112. doi:10.1006/bijl.1999.0332

Hewitt, G.M., Prendini, L., Savary, W.E., et al., 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biol. J. Linn. Soc. 58, 247–276. doi:10.1111/j.1095-8312.1996.tb01434.x

Hopkin, S.P., 1997. Biology of the springtails: (Insecta: Collembola). Oxford University Press, Oxford, UK.

Kotlík, P., Deffontaine, V., Mascheretti, S., Zima, J., Michaux, J.R., Searle, J.B., 2006. A northern glacial refugium for bank voles (*Clethrionomys glareolus*). Proc. Natl. Acad. Sci. 103, 14860–14864. doi:10.1073/pnas.0603237103

Kováč, Ľ., Parimuchová, A., Miklisová, D., 2016. Distributional patterns of cave Collembola (Hexapoda) in association with habitat conditions, geography and subterranean refugia in the Western Carpathians. Biol. J. Linn. Soc. 119, 571–592. doi:10.1111/bij.12555

Krijgsman, W., Hilgen, F.J., Raffi, I., Sierro, F.J., Wilson, D.S., 1999. Chronology, causes and progression of the Messinian salinity crisis. Nature 400, 652–655. doi:10.1038/23231

Leigh, J.W., Bryant, D., 2015. POPART:full-feature software for haplotype network construction. Methods Ecol. Evol. 6, 1110–1116. doi:10.1111/2041-210X.12410

Leinaas, H.P., 1981. Activity of Arthropoda in snow within a coniferous forest, with special reference to Collembola. Ecography (Cop.). 4, 127–138. doi:10.1111/j.1600-0587.1981.tb00989.x

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452. doi:10.1093/bioinformatics/btp187

Louwye, S., Foubert, A., Mertens, K., Van Rooij, D., 2008. Integrated stratigraphy and palaeoecology of the Lower and Middle Miocene of the Porcupine Basin. Geol. Mag. 145, 321–344. doi:10.1017/S0016756807004244

Mateos, E., 2012. The European *Lepidocyrtus lanuginosus* group (Collembola: Entomobryidae), definition and description of a new species from Spain. Zootaxa 3570, 69–81.

Mateos, E., Escuer, P., Buşmachiu, G., Riutort, M., Álvarez-Presas, M., 2018. Untangling *Lepidocyrtus* (Collembola, Entomobryidae): new molecular data shed light on the relationships of the European groups. Invertebr. Syst. 32, 639–651. doi:10.1071/IS17056

McGaughran, A., Stevens, M.I., Holland, B.R., 2010. Biogeography of circum-Antarctic springtails. Mol. Phylogenet. Evol. 57, 48–58. doi:10.1016/j.ympev.2010.06.003

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, in: 2010 Gateway Computing Environments Workshop (GCE). IEEE, pp. 1–8. doi:10.1109/GCE.2010.5676129

Mittelbach, G.G., Schemske, D.W., Cornell, H. V., Allen, A.P., Brown, J.M., Bush, M.B., Harrison, S.P., Hurlbert, A.H., Knowlton, N., Lessios, H.A., McCain, C.M., McCune, A.R., McDade, L.A., McPeek, M.A., Near, T.J., Price, T.D., Ricklefs, R.E., Roy, K., Sax, D.F., Schluter, D., Sobel, J.M., Turelli, M., 2007. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. Ecol. Lett. 10, 315–331. doi:10.1111/j.1461-0248.2007.01020.x

Moore, P.D., 2002. Springboards for sprintails. Nature 418, 381.

Mosbrugger, V., Utescher, T., Dilcher, D.L., 2005. Cenozoic continental climatic evolution of Central Europe. Proc. Natl. Acad. Sci. 102, 14964–14969. doi:10.1073/pnas.0505267102

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. doi:10.1093/molbev/msu300

Normand, S., Ricklefs, R.E., Skov, F., Bladt, J., Tackenberg, O., Svenning, J.-C., 2011. Postglacial migration supplements climate in determining plant species ranges in Europe. Proc. R. Soc. B Biol. Sci. 278, 3644–3653. doi:10.1098/rspb.2010.2769

Papadopoulou, A., Anastasiou, I., Vogler, A.P., 2010. Revisiting the insect mitochondrial molecular clock: The mid-aegean trench calibration. Mol. Biol. Evol. 27, 1659–1672. doi:10.1093/molbev/msq051

Porco, D., Bedos, A., Greenslade, P., Janion, C., Skarzyński, D., Stevens, M.I., Jansen van Vuuren, B., Deharveng, L., 2012a. Challenging species delimitation in Collembola: cryptic diversity among common springtails unveiled by DNA barcoding. Invertebr. Syst. 26, 470–477. doi:10.1071/IS12026

Porco, D., Decaëns, T., Deharveng, L., James, S.W., Skarzyński, D., Erséus, C., Butt, K.R., Richard, B., Hebert, P.D.N., 2013. Biological invasions in soil: DNA barcoding as a monitoring tool in a multiple taxa survey targeting European earthworms and springtails in North America. Biol. Invasions 15, 899–910. doi:10.1007/s10530-012-0338-2

Porco, D., Potapov, M., Bedos, A., Busmachiu, G., Weiner, W.M., Hamra-Kroua, S., Deharveng, L., 2012b. Cryptic diversity in the ubiquist species *Parisotoma notabilis* (Collembola, Isotomidae): a long-used chimeric species? PLoS One 7, e46056. doi:10.1371/journal.pone.0046056

Porco, D., Skarzyński, D., Decaëns, T., Hebert, P.D.N., Deharveng, L., 2014. Barcoding the Collembola of Churchill: A molecular taxonomic reassessment of species diversity in a sub-Arctic area. Mol. Ecol. Resour. 14, 249–261. doi:10.1111/1755-0998.12172

Provan, J., Bennett, K.D., 2008. Phylogeographic insights into cryptic glacial refugia. Trends Ecol. Evol. 23, 564–571. doi:10.1016/j.tree.2008.06.010

Rambaut, 2012. FigTree 1.4.3 [WWW Document]. URL http://tree.bio.ed.ac.uk/software/figtree/ (accessed 8.30.18).

Rambaut, A., Drummond, A.J., M., S., 2013. Tracer v1.6 [WWW Document]. URL http://beast.bio.ed.ac.uk/Tracer (accessed 8.30.18).

Salmon, S., Ponge, J.F., Gachet, S., Deharveng, L., Lefebvre, N., Delabrosse, F., 2014. Linking species, traits and habitat characteristics of Collembola at European scale. Soil Biol. Biochem. 75, 73–85. doi:10.1016/j.soilbio.2014.04.002

Shapiro, B., Drummond, A.J., Rambaut, A., Wilson, M.C., Matheus, P.E., Sher, A. V., Pybus, O.G., Gilbert, M.T.P., Barnes, I., Binladen, J., Willerslev, E., Hansen, A.J., Baryshnikov, G.F., Burns, J.A., Davydov, S., Driver, J.C., Froese, D.G., Harington, C.R., Keddie, G., Kosintsev, P., Kunz, M.L., Martin, L.D., Stephenson, R.O., Storer, J., Tedford, R., Zimov, S., Cooper, A., 2004. Rise and fall of the Beringian steppe bison. Science (80-.). 306, 1561–1565. doi:10.1126/science.1101074

Sommer, R.S., Nadachowski, A., 2006. Glacial refugia of mammals in Europe: Evidence from fossil records. Mamm. Rev. doi:10.1111/j.1365-2907.2006.00093.x

Stewart, J.R., Dalén, L., 2008. Is the glacial refugium concept relevant for northern species? A comment on Pruett and Winker 2005. Clim. Change 86, 19–22. doi:10.1007/s10584-007-9366-9

Stewart, J.R., Lister, A.M., Barnes, I., Dalen, L., 2010. Refugia revisited: individualistic responses of species in space and time. Proc. R. Soc. B Biol. Sci. 277, 661–671. doi:10.1098/rspb.2009.1272

Sun, X., Zhang, F., Ding, Y., Davies, T.W., Li, Y., Wu, D., 2017. Delimiting species of *Protaphorura* (Collembola: Onychiuridae): integrative evidence based on morphology, DNA sequences and geography. Sci. Rep. 7, 8261. doi:10.1038/s41598-017-08381-4

Taberlet, P., Fumagalli, L., Wust-Saucy, A.G., Cosson, J.F., 1998. Comparative phylogeography and postglacial colonization routes in Europe. Mol. Ecol. 7, 453–464. doi:10.1046/j.1365-294x.1998.00289.x

Teets, N.M., Denlinger, D.L., 2014. Surviving in a frozen desert: environmental stress physiology of terrestrial Antarctic arthropods. J. Exp. Biol. 217, 84–93. doi:10.1242/jeb.089490

Timmermans, M.J.T.N., Ellers, J., Mariën, J., Verhoef, S.C., Ferwerda, E.B., van Straalen, N.M., 2005. Genetic structure in Orchesella cincta (Collembola): Strong subdivision of European populations inferred from mtDNA and AFLP markers. Mol. Ecol. 14, 2017–2024. doi:10.1111/j.1365-294X.2005.02548.x

Ulrich, W., Fiera, C., 2009. Environmental correlates of species richness of European springtails (Hexapoda: Collembola). Acta Oecologica 35, 45–52. doi:10.1016/j.actao.2008.07.007

Ursenbacher, S., Carlsson, M., Helfer, V., Tegelström, H., Fumagalli, L., 2006. Phylogeography and Pleistocene refugia of the adder (*Vipera berus*) as inferred from mitochondrial DNA sequence data. Mol. Ecol. 15, 3425–3437. doi:10.1111/j.1365-294X.2006.03031.x

Vaidya, G., Lohman, D.J., Meier, R., 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27, 171–180. doi:10.1111/j.1096-0031.2010.00329.x

van der Wurff, A.W.G., Gols, R., Ernsting, G., van Straalen, N.M., 2005. Population genetic structure of *Orchesella cincta* (Collembola; Hexapoda) in NW Europe, as revealed by microsatellite markers. Pedobiologia. 49, 167–174. doi:10.1016/j.pedobi.2004.10.004

Varga, Z.S., Schmitt, T., 2008. Types of oreal and oreotundral disjunctions in the western Palearctic. Biol. J. Linn. Soc. 93, 415–430. doi:10.1111/j.1095-8312.2007.00934.x

von Saltzwedel, H., Scheu, S., Schaefer, I., 2017. Genetic structure and distribution of *Parisotoma notabilis* (Collembola) in Europe: Cryptic diversity, split of lineages and colonization patterns. PLoS One 12, e0170909. doi:10.1371/journal.pone.0170909

von Saltzwedel, H., Scheu, S., Schaefer, I., 2016. Founder events and pre-glacial divergences shape the genetic structure of European Collembola species. BMC Evol. Biol. 16, 148. doi:10.1186/s12862-016-0719-8

Zachos, J., Pagani, H., Sloan, L., Thomas, E., Billups, K., 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. Science (80-.). 292, 686–693. doi:10.1126/science.1059412

Zhang, B., Chang, L., Ni, Z., Callaham, M.A., Sun, X., Wu, D., 2014. Effects of land use changes on winter-active Collembola in Sanjiang Plain of China. Appl. Soil Ecol. 83, 51–58. doi:10.1016/j.apsoil.2014.03.008

Zhang, B., Chang, L., Ni, Z., Sun, X., Wu, D., 2017. Directional migration of three *Desoria species* (Collembola: Isotomidae) on the snow surface in late winter. Eur. J. Soil Biol. 81, 64–68. doi:10.1016/j.ejsobi.2017.06.009

Zhang, B., Chen, T.W., Mateos, E., Scheu, S., Schaefer, I., 2018. Cryptic species in *Lepidocyrtus lanuginosus* (Collembola: Entomobryidae) are sorted by habitat type. Pedobiologia. 68, 12–19. doi:10.1016/j.pedobi.2018.03.001

Zhang, F., Jantarit, S., Nilsai, A., Stevens, M.I., Ding, Y., Satasook, C., 2018. Species delimitation in the morphologically conserved *Coecobrya* (Collembola: Entomobryidae): A case study integrating morphology and molecular traits to advance current taxonomy. Zool. Scr. 47, 342–356. doi:10.1111/zsc.12279

Supplementary Materials

Supplementary Table S4.1 Sampling data of Collembola specimens used for DNA barcoding. 28SD1–2, COII, COI and elongation factor 1-α accessions represents the registration numbers of sequences in GenBank.

Species	Individual name	Locality	Habitat	Position	Collecting date	28SD1_2	COII	COI	EF1-α
Lepidocyrtus bicoloris	LEbicoloris_Tag_1	Tagamanent	Grassland surround by forest	41°45'3.68"N, 2°18'20.50"E	15/10/2011	MH570534	MH570690		MH570984
Lepidocyrtus bicoloris	LEbicoloris_Tag_2					MH570535	MH570691		MH570985
Lepidocyrtus bicoloris	LEbicoloris_Tag_3					MH570536	MH570692	MH570854	MH570986
Lepidocyrtus bicoloris	LEbicoloris_Tag_4					MH570537	MH570693		
Lepidocyrtus bicoloris	LEbicoloris_Tag_5					MH570538	MH570694	MH570855	
Lepidocyrtus cyaneus	LEcyaneus_Ami_1	Amiens	Grassland next to forest	42°44'33.41"N, 1°46'59.04"E	11/05/2015	MH570544	MH570700		
Lepidocyrtus cyaneus	LEcyaneus_Ami_2					MH570545	MH570701	MH570861	MH570992
Lepidocyrtus cyaneus	LEcyaneus_Ami_3					MH570546	MH570702	MH570862	MH570993
Lepidocyrtus cyaneus	LEcyaneus_Ami_4					MH570547	MH570703		MH570994
Lepidocyrtus cyaneus	LEcyaneus_Ami_5					MH570548	MH570704	MH570863	MH570995
Lepidocyrtus cyaneus	LEcyaneus_Ant_1	Antwerpen	Grassland-arable land	49°50'33.9"N, 2°31'35.7"E	18/04/2016	MH570549	MH570705		MH570996
Lepidocyrtus cyaneus	LEcyaneus_Ant_2					MH570550	MH570706	MH570864	MH570997
Lepidocyrtus cyaneus	LEcyaneus_Ant_3					MH570551	MH570707	MH570865	MH570998
Lepidocyrtus cyaneus	LEcyaneus_Ant_4					MH570552	MH570708	MH570866	MH570999
Lepidocyrtus cyaneus	LEcyaneus_Ant_5					MH570553	MH570709	MH570867	MH571000
Lepidocyrtus cyaneus	LEcyaneus_Bre_1	Bremen North	Grassland next to forest	53°14'36.1"N, 8°39'20.1"E	19/04/2016	MH570554	MH570710	MH570868	MH571001
Lepidocyrtus cyaneus	LEcyaneus_Bre_2					MH570555	MH570711	MH570869	MH571002
Lepidocyrtus cyaneus	LEcyaneus_Bre_3						MH570712	MH570870	MH571003
Lepidocyrtus cyaneus	LEcyaneus_Bre_4					MH570556	MH570713	MH570871	MH571004
Lepidocyrtus cyaneus	LEcyaneus_Bre_5						MH570714	MH570872	MH571005
Lepidocyrtus cyaneus	LEcyaneus_Cas_1	Cassine	Grassland surround by forest	44°46'8.09"N, 8°29'28.64"E	07/05/2015	MH570557	MH570715	MH570873	MH571006
Lepidocyrtus cyaneus	LEcyaneus_Cas_2					MH570558	MH570716	MH570874	MH571007
Lepidocyrtus cyaneus	LEcyaneus_Cas_3					MH570559	MH570717	MH570875	MH571008

Species	Individual name	Locality	Habitat	Position	Collecting date	28SD1_2	COII	COI	EF1-α
Lepidocyrtus cyaneus	LEcyaneus_Cas_4					MH570560	MH570718		
Lepidocyrtus cyaneus	LEcyaneus_Cas_5					MH570561	MH570719	MH570876	MH571009
Lepidocyrtus cyaneus	LEcyaneus_Cth_1	Chateau-Thierry	Grassland-arable land	49° 5'45.17"N, 3°26'12.35"E	13/05/2015	MH570562	MH570720		MH571010
Lepidocyrtus cyaneus	LEcyaneus_Cth_2					MH570563	MH570721	MH570877	MH571011
Lepidocyrtus cyaneus	LEcyaneus_Cth_3					MH570564	MH570722		MH571012
Lepidocyrtus cyaneus	LEcyaneus_Cth_4					MH570565	MH570723	MH570878	MH571013
Lepidocyrtus cyaneus	LEcyaneus_Cth_5					MH570566	MH570724	MH570879	MH571014
Lepidocyrtus cyaneus	LEcyaneus_Ill_1	Illertissen	Grassland next to forest	48°13'41.43"N, 10° 7'1.12"E	03/05/2015	MH570567	MH570725	MH570880	
Lepidocyrtus cyaneus	LEcyaneus_Ill_2					MH570568	MH570726	MH570881	MH571015
Lepidocyrtus cyaneus	LEcyaneus_Ill_3					MH570569	MH570727	MH570882	MH571016
Lepidocyrtus cyaneus	LEcyaneus_Ill_4					MH570570	MH570728	MH570883	MH571017
Lepidocyrtus cyaneus	LEcyaneus_Ill_5					MH570571	MH570729	MH570884	
Lepidocyrtus cyaneus	LEcyaneus_Ill_6					MH570572	MH570730	MH570885	
Lepidocyrtus cyaneus	LEcyaneus_Jen_1	Jena	Grassland-arable land	50°57'0.96"N, 11°37'16.43"E	12/05/2016	MH570573	MH570731	MH570886	MH571018
Lepidocyrtus cyaneus	LEcyaneus_Jen_2					MH570574	MH570732	MH570887	
Lepidocyrtus cyaneus	LEcyaneus_Jen_3					MH570575	MH570733	MH570888	MH571019
Lepidocyrtus cyaneus	LEcyaneus_Jen_4					MH570576	MH570734	MH570889	MH571020
Lepidocyrtus cyaneus	LEcyaneus_Jen_5					MH570577	MH570735	MH570890	MH571021
Lepidocyrtus cyaneus	LEcyaneus_Lei_1	Leipzig	Grassland surround by forest	51°19'33.78"N, 12°21'21.05"E	25/04/2016	MH570578	MH570736	MH570891	MH571022
Lepidocyrtus cyaneus	LEcyaneus_Lei_2					MH570579	MH570737	MH570892	MH571023
Lepidocyrtus cyaneus	LEcyaneus_Lei_3					MH570580	MH570738	MH570893	MH571024
Lepidocyrtus cyaneus	LEcyaneus_Lei_4					MH570581	MH570739	MH570894	MH571025
Lepidocyrtus cyaneus	LEcyaneus_Lei_5					MH570582	MH570740		
Lepidocyrtus cyaneus	LEcyaneus_Ossa_1	Ossenfeld	Arable land	51°32'52.4"N, 9°47'52.9"E	17/04/2014	MH160158	MH144023	MH143938	MH178011
Lepidocyrtus cyaneus	LEcyaneus_Ossg_1	Ossenfeld	Grassland next to forest	51°32'50.3"N, 9°47'50.4"E	17/04/2014	MH160159	MH144024	MH143939	MH178012
Lepidocyrtus cyaneus	LEcyaneus_Ossg_2					MH160160	MH144025	MH143940	MH178013
Lepidocyrtus cyaneus	LEcyaneus_Waaa_1	Waake	Arable land		19/06/2014	MH160161	MH144026	MH143941	MH178014

Species	Individual name	Locality	Habitat	Position	Collecting date	28SD1_2	COII	COI	EF1-α
Lepidocyrtus cyaneus	LEcyaneus_Waaa_2			51°33'47.1"N, 10°03'30.4"E		MH160162	MH144027	MH143942	MH178015
Lepidocyrtus cyaneus	LEcyaneus_Waaa_3					MH160163	MH144028	MH143943	MH178016
Lepidocyrtus cyaneus	LEcyaneus_Mal_1	Marseille-Allauch	Grassland surround by forest	43°22'16.57"N, 5°32'9.46"E	09/05/2015	MH570583	MH570741	MH570895	MH571026
Lepidocyrtus cyaneus	LEcyaneus_Mal_2					MH570584	MH570742	MH570896	MH571027
Lepidocyrtus cyaneus	LEcyaneus_Mal_3					MH570585	MH570743	MH570897	MH571028
Lepidocyrtus cyaneus	LEcyaneus_Mal_4					MH570586	MH570744		MH571029
Lepidocyrtus cyaneus	LEcyaneus_Mal_5					MH570587	MH570745		
Lepidocyrtus cyaneus	LEcyaneus_Oyp_1	Oye-Plage (Calais)	Grassland next to forest	50°59'47.5"N, 2°02'31.9"E	18/04/2016	MH570588	MH570746	MH570898	
Lepidocyrtus cyaneus	LEcyaneus_Oyp_2					MH570589	MH570747	MH570899	
Lepidocyrtus cyaneus	LEcyaneus_Oyp_3					MH570590	MH570748	MH570900	
Lepidocyrtus cyaneus	LEcyaneus_Oyp_4					MH570591	MH570749		
Lepidocyrtus cyaneus	LEcyaneus_Oyp_5					MH570592	MH570750		
Lepidocyrtus cyaneus	LEcyaneus_Pin_1	"Ordesa y Monte Perdido" National Park, Pineta valley	Grassland surround by forest	42°40'35.3"N, 0°05'07.5"E	30/05/2009	MH570593	MH570751	MH570901	MH571030
Lepidocyrtus cyaneus	LEcyaneus_Pin_2					MH570594	MH570752	MH570902	MH571031
Lepidocyrtus cyaneus	LEcyaneus_Pin_3					MH570595	MH570753		MH571032
Lepidocyrtus cyaneus	LEcyaneus_Pin_4					MH570596	MH570754		MH571033
Lepidocyrtus cyaneus	LEcyaneus_Pin_5					MH570597	MH570755		MH571034
Lepidocyrtus cyaneus	LEcyaneus_Rap_1	Rapallo	Grassland surround by forest	44°22'48.53"N, 9°30'40.63"E	06/05/2015	MH570598	MH570756	MH570903	
Lepidocyrtus cyaneus	LEcyaneus_Rap_2					MH570599	MH570757	MH570904	MH571035
Lepidocyrtus cyaneus	LEcyaneus_Rap_3					MH570600	MH570758		
Lepidocyrtus cyaneus	LEcyaneus_Rap_4					MH570601	MH570759	MH570905	
Lepidocyrtus cyaneus	LEcyaneus_Rap_5					MH570602	MH570760	MH570906	MH571036
Lepidocyrtus cyaneus	LEcyaneus_Sch_1	Schwechat (Wien Airport)	Grassland next to forest	48°08'38.0"N, 16°31'32.1"E	23/04/2016	MH570603	MH570761	MH570907	MH571037
Lepidocyrtus cyaneus	LEcyaneus_Sch_2					MH570604	MH570762	MH570908	MH571038
Lepidocyrtus cyaneus	LEcyaneus_Sch_3					MH570605	MH570763	MH570909	MH571039
Lepidocyrtus cyaneus	LEcyaneus_Sch_4					MH570606	MH570764	MH570910	MH571040
Lepidocyrtus cyaneus	LEcyaneus_Sch_5					MH570607	MH570765	MH570911	MH571041
Lepidocyrtus cyaneus	LEcyaneus_Smen_1				14/05/2015	MH570608	MH570766	MH570912	MH571042

Species	Individual name	Locality	Habitat	Position	Collecting date	28SD1_2	COII	COI	EF1-α
Lepidocyrtus cyaneus	LEcyaneus_Smen_2	Sainte-Menehoult	Grassland-arable land	49° 5'46.17"N, 4°55'26.67"E		MH570609	MH570767	MH570913	MH571043
Lepidocyrtus cyaneus	LEcyaneus_Smen_3					MH570610	MH570768	MH570914	MH571044
Lepidocyrtus cyaneus	LEcyaneus_Smen_4					MH570611	MH570769	MH570915	MH571045
Lepidocyrtus cyaneus	LEcyaneus_Smen_5					MH570612	MH570770	MH570916	MH571046
Lepidocyrtus cyaneus	LEcyaneus_Utr_1	Utrecht	Grassland surround by forest	52°08'50.6"N, 5°12'21.9"E	19/04/2016	MH570613	MH570771	MH570917	MH571047
Lepidocyrtus cyaneus	LEcyaneus_Utr_2					MH570614	MH570772	MH570918	MH571048
Lepidocyrtus cyaneus	LEcyaneus_Utr_3					MH570615	MH570773	MH570919	MH571049
Lepidocyrtus cyaneus	LEcyaneus_Utr_4						MH570774		MH571050
Lepidocyrtus cyaneus	LEcyaneus_Utr_5					MH570616	MH570775	MH570920	MH571051
Lepidocyrtus cyaneus	LEcyaneus_Ver_1	Vernante	Grassland surround by forest	44°12'6.99"N, 7°30'17.58"E	07/05/2015	MH570617	MH570776	MH570921	MH571052
Lepidocyrtus cyaneus	LEcyaneus_Ver_2					MH570618	MH570777	MH570922	MH571053
Lepidocyrtus cyaneus	LEcyaneus_Ver_3					MH570619	MH570778	MH570923	
Lepidocyrtus cyaneus	LEcyaneus_Ver_4						MH570779		MH571054
Lepidocyrtus cyaneus	LEcyaneus_Ver_5					MH570620	MH570780	MH570924	
Lepidocyrtus lanuginosus	LElanuginosus_Aig_1	"Montseny" mountain (Aiguafreda)	Forest	41°46'13.8"N, 2°16'19.9"E	10/02/2007	MH570621	MH570781	MH570925	MH571055
Lepidocyrtus lanuginosus	LElanuginosus_Aig_2					MH570622	MH570782	MH570926	MH571056
Lepidocyrtus lanuginosus	LElanuginosus_Aig_4					MH570623	MH570783	MH570927	MH571057
Lepidocyrtus lanuginosus	LElanuginosus_Aig_5					MH570624	MH570784	MH570928	MH571058
Lepidocyrtus lanuginosus	LElanuginosus_Alt_1	Altmoorhausen	Grassland-arable land	53° 3'59.74"N, 8°21'1.83"E	18/05/2016	MH570625	MH570785	MH570929	MH571059
Lepidocyrtus lanuginosus	LElanuginosus_Alt_2					MH570626	MH570786	MH570930	MH571060
Lepidocyrtus lanuginosus	LElanuginosus_Alt_3					MH570627	MH570787	MH570931	MH571061
Lepidocyrtus lanuginosus	LElanuginosus_Alt_4					MH570628	MH570788	MH570932	MH571062
Lepidocyrtus lanuginosus	LElanuginosus_Alt_5					MH570629	MH570789	MH570933	MH571063
Lepidocyrtus lanuginosus	LElanuginosus_Bila_1	Billingshausen arable land	Arable land	51°35'24.2"N, 10°01'35.6"E	30/04/2014	MH160146	MH144011		
Lepidocyrtus lanuginosus	LElanuginosus_Bila_2					MH160147	MH144012		
Lepidocyrtus lanuginosus	LElanuginosus_Bila_3					MH160148	MH144013	MH143935	MH178008
Lepidocyrtus lanuginosus	LElanuginosus_Bilg_1	Billingshausen grassland		51°35'37.4"N, 10°01'51.6"E	30/04/2014	MH160154	MH144019	MH143937	MH178010
Lepidocyrtus lanuginosus	LElanuginosus_Bilg_2					MH160155	MH144020		

Species	Individual name	Locality	Habitat	Position	Collecting date	28SD1_2	COII	COI	EF1-α
Lepidocyrtus lanuginosus	LElanuginosus_Bilg_3		Grassland next to forest			MH160156	MH144021		
Lepidocyrtus lanuginosus	LElanuginosus_Bilg_4					MH160157	MH144022		
Lepidocyrtus lanuginosus	LElanuginosus_Bilf_1	Billingshausen forest	Forest	51°35'32.5"N, 10°01'57.1"E	30/04/2014	MH160149	MH144014	MH143936	MH178009
Lepidocyrtus lanuginosus	LElanuginosus_Bilf_2					MH160150	MH144015		
Lepidocyrtus lanuginosus	LElanuginosus_Bilf_3					MH160151	MH144016		
Lepidocyrtus lanuginosus	LElanuginosus_Bilf_4					MH160152	MH144017		
Lepidocyrtus lanuginosus	LElanuginosus_Bilf_5					MH160153	MH144018		
Lepidocyrtus lanuginosus	LElanuginosus_Brn_1	Brno	Grassland next to forest	49°11'32.15"N, 16°25'33.73"E	24/04/2016	MH570630	MH570790		
Lepidocyrtus lanuginosus	LElanuginosus_Brn_2						MH570791		
Lepidocyrtus lanuginosus	LElanuginosus_Brn_3					MH570631	MH570792		
Lepidocyrtus lanuginosus	LElanuginosus_Brn_4						MH570793	MH570934	
Lepidocyrtus lanuginosus	LElanuginosus_Brn_5						MH570794	MH570935	MH571064
Lepidocyrtus lanuginosus	LElanuginosus_Cat_1	"Serra de Catllaràs" mountain	Forest	42°13'44.0"N, 1°56'16.8"E	05/04/2008	MH570632	MH570795	MH570936	MH571065
Lepidocyrtus lanuginosus	LElanuginosus_Cat_2					MH570633	MH570796	MH570937	MH571066
Lepidocyrtus lanuginosus	LElanuginosus_Cat_3					MH570634	MH570797	MH570938	MH571067
Lepidocyrtus lanuginosus	LElanuginosus_Cat_4					MH570635	MH570798	MH570939	MH571068
Lepidocyrtus lanuginosus	LElanuginosus_Cat_5					MH570636	MH570799	MH570940	MH571069
Lepidocyrtus lanuginosus	LElanuginosus_Depa_3	Deppoldshausen arable land	Arable land	51°34'32.3"N, 9°58'23.2"E	27/03/2014	MH160114	MH143979	MH143927	MH177999
Lepidocyrtus lanuginosus	LElanuginosus_Depa_4					MH160115	MH143980		
Lepidocyrtus lanuginosus	LElanuginosus_Depa_5					MH160116	MH143981		
Lepidocyrtus lanuginosus	LElanuginosus_Depa_6					MH160117	MH143982		
Lepidocyrtus lanuginosus	LElanuginosus_Depa_7					MH160118	MH143983		
Lepidocyrtus lanuginosus	LElanuginosus_Depf_1	Deppoldshausen forest	Forest	51°34'30.2"N, 9°58'28.0"E	27/03/2014	MH160119	MH143984		
Lepidocyrtus lanuginosus	LElanuginosus_Depf_2					MH160120	MH143985		
Lepidocyrtus lanuginosus	LElanuginosus_Depf_3					MH160121	MH143986		
Lepidocyrtus lanuginosus	LElanuginosus_Depf_4					MH160122	MH143987	MH143928	MH178002
Lepidocyrtus lanuginosus	LElanuginosus_Depf_5					MH160123	MH143988	MH143929	MH178003

Species	Individual name	Locality	Habitat	Position	Collecting date	28SD1_2	COII	COI	EF1-α
Lepidocyrtus lanuginosus	LElanuginosus_Ella_1	Ellershausen arable land	Arable land	51°30'45.5"N, 9°40'05.9"E	14/04/2014	MH160124	MH143989		
Lepidocyrtus lanuginosus	LElanuginosus_Ella_2					MH160125	MH143990		
Lepidocyrtus lanuginosus	LElanuginosus_Ella_3					MH160126	MH143991		
Lepidocyrtus lanuginosus	LElanuginosus_Ellf_1	Ellershausen forest	Forest	51°30'47.7"N, 9°39'58.6"E	14/04/2014	MH160127	MH143992		
Lepidocyrtus lanuginosus	LElanuginosus_Ellf_2					MH160128	MH143993		
Lepidocyrtus lanuginosus	LElanuginosus_Ellf_3					MH160129	MH143994		
Lepidocyrtus lanuginosus	LElanuginosus_Ellf_4					MH160130	MH143995		
Lepidocyrtus lanuginosus	LElanuginosus_Ellf_5					MH160131	MH143996		
Lepidocyrtus lanuginosus	LElanuginosus_Hera_1	Herberhausen arable land	Arable land	51°32'02.7"N, 10°00'02.1"E	19/06/2014	MH160100	MH143965	MH143920	MH177993
Lepidocyrtus lanuginosus	LElanuginosus_Hera_2					MH160101	MH143966	MH143921	MH177994
Lepidocyrtus lanuginosus	LElanuginosus_Hera_3					MH160102	MH143967		
Lepidocyrtus lanuginosus	LElanuginosus_Hera_4					MH160103	MH143968		
Lepidocyrtus lanuginosus	LElanuginosus_Herg_2	Herberhausen grassland	Grassland next to forest	51°31'58.6"N, 9°59'33.8"E	19/06/2014	MH160109	MH143974	MH143924	MH177996
Lepidocyrtus lanuginosus	LElanuginosus_Herg_3					MH160110	MH143975	MH143925	MH177997
Lepidocyrtus lanuginosus	LElanuginosus_Herg_4					MH160111	MH143976	MH143926	MH177998
Lepidocyrtus lanuginosus	LElanuginosus_Herg_5					MH160112	MH143977		
Lepidocyrtus lanuginosus	LElanuginosus_Herg_6					MH160113	MH143978	MH143923	MH177995
Lepidocyrtus lanuginosus	LElanuginosus_Herf_3	Herberhausen forest	Forest	51°31'50.6"N, 9°59'26.4"E	19/06/2014	MH160104	MH143969		
Lepidocyrtus lanuginosus	LElanuginosus_Herf_4					MH160105	MH143970		
Lepidocyrtus lanuginosus	LElanuginosus_Herf_5					MH160106	MH143971		
Lepidocyrtus lanuginosus	LElanuginosus_Herf_6					MH160107	MH143972	MH143922	MH178004
Lepidocyrtus lanuginosus	LElanuginosus_Herf_7					MH160108	MH143973		
Lepidocyrtus lanuginosus	LElanuginosus_Hor_1	Horka (Jungbunzlau)	Grassland-arable land	50°28'40.2"N, 14°58'43.0"E	24/04/2016	MH570637	MH570800	MH570941	MH571070
Lepidocyrtus lanuginosus	LElanuginosus_Hor_2					MH570638	MH570801		MH571071
Lepidocyrtus lanuginosus	LElanuginosus_Hor_3					MH570639	MH570802	MH570942	MH571072
Lepidocyrtus lanuginosus	LElanuginosus_Hor_4					MH570640	MH570803	MH570943	
Lepidocyrtus lanuginosus	LElanuginosus_Hor_5					MH570641	MH570804		MH571073
Lepidocyrtus lanuginosus	LElanuginosus_Jen_1	Jena			12/05/2016	MH570642	MH570805		MH571074

Species	Individual name	Locality	Habitat	Position	Collecting date	28SD1_2	COII	COI	EF1-α
Lepidocyrtus lanuginosus	LElanuginosus_Jen_2		Grassland-arable land	50°57'0.96"N, 11°37'16.43"E		MH570643	MH570806		MH571075
Lepidocyrtus lanuginosus	LElanuginosus_Jen_3					MH570644	MH570807	MH570944	MH571076
Lepidocyrtus lanuginosus	LElanuginosus_Jen_4					MH570645		MH570945	MH571077
Lepidocyrtus lanuginosus	LElanuginosus_Kut_1	Kutno	Grassland next to forest	52°16'38.8"N, 19°20'54.4"E	21/04/2016	MH570646	MH570808	MH570946	MH571078
Lepidocyrtus lanuginosus	LElanuginosus_Kut_2					MH570647	MH570809	MH570947	MH571079
Lepidocyrtus lanuginosus	LElanuginosus_Kut_3					MH570648	MH570810	MH570948	MH571080
Lepidocyrtus lanuginosus	LElanuginosus_Kut_4						MH570811	MH570949	MH571081
Lepidocyrtus lanuginosus	LElanuginosus_Kut_5					MH570649	MH570812	MH570950	MH571082
Lepidocyrtus lanuginosus	LElanuginosus_Lebe_1	Lebeny west	Grassland-arable land	47°44'55.8"N, 17°21'08.7"E	23/04/2016		MH570813	MH570951	
Lepidocyrtus lanuginosus	LElanuginosus_Lebe_2						MH570814	MH570952	MH571083
Lepidocyrtus lanuginosus	LElanuginosus_Lebe_3					MH570650	MH570815	MH570953	MH571084
Lepidocyrtus lanuginosus	LElanuginosus_Lebe_4					MH570651	MH570816	MH570954	MH571085
Lepidocyrtus lanuginosus	LElanuginosus_Mon_1	"Montnegre" mountain	Forest	41°39'52.2"N, 2°33'46.0"E	18/04/2007	MH570652	MH570817	MH570955	MH571086
Lepidocyrtus lanuginosus	LElanuginosus_Mon_2					MH570653	MH570818	MH570956	MH571087
Lepidocyrtus lanuginosus	LElanuginosus_Mon_3					MH570654	MH570819	MH570957	MH571088
Lepidocyrtus lanuginosus	LElanuginosus_Mon_4					MH570655	MH570820	MH570958	
Lepidocyrtus lanuginosus	LElanuginosus_Mon_5					MH570656	MH570821	MH570959	MH571089
Lepidocyrtus lanuginosus	LElanuginosus_Niea_10	Nitra East	Grassland-arable land	48°18'13.9"N, 18°09'11.8"E	23/04/2016	MH570657	MH570822	MH570960	MH571090
Lepidocyrtus lanuginosus	LElanuginosus_Niea_6					MH570658	MH570823	MH570961	MH571091
Lepidocyrtus lanuginosus	LElanuginosus_Niea_7					MH570659	MH570824	MH570962	MH571092
Lepidocyrtus lanuginosus	LElanuginosus_Niea_8					MH570660	MH570825	MH570963	MH571093
Lepidocyrtus lanuginosus	LElanuginosus_Niea_9					MH570661	MH570826	MH570964	MH571094
Lepidocyrtus lanuginosus	LElanuginosus_Ossa_1	Ossenfeld arable land	Forest	51°32'52.4"N, 9°47'52.9"E	14/04/2014	MH160132	MH143997		
Lepidocyrtus lanuginosus	LElanuginosus_Ossf_1	Ossenfeld forest	Forest	51°32'56.3"N, 9°48'01.5"E	14/04/2014	MH160133	MH143998	MH143930	MH178005
Lepidocyrtus lanuginosus	LElanuginosus_Ossf_2					MH160134	MH143999		
Lepidocyrtus lanuginosus	LElanuginosus_Ossf_3					MH160135	MH144000		
Lepidocyrtus lanuginosus	LElanuginosus_Ott_1	Otterbach			14/05/2015	MH570662	MH570827	MH570965	MH571095

Species	Individual name	Locality	Habitat	Position	Collecting date	28SD1_2	COII	COI	EF1-α
Lepidocyrtus lanuginosus	LElanuginosus_Ott_2		Grassland surround by forest	49°26'31.74"N, 7°55'19.24"E		MH570663	MH570828	MH570966	MH571096
Lepidocyrtus lanuginosus	LElanuginosus_Ott_3					MH570664	MH570829	MH570967	MH571097
Lepidocyrtus lanuginosus	LElanuginosus_Ott_4					MH570665	MH570830		MH571098
Lepidocyrtus lanuginosus	LElanuginosus_Ott_5					MH570666	MH570831	MH570968	MH571099
Lepidocyrtus lanuginosus	LElanuginosus_Pga_1	Park Gardasee	Grassland surround by forest	45°42'15.73"N, 10°58'23.94"E	04/05/2015	MH570667	MH570832		
Lepidocyrtus lanuginosus	LElanuginosus_Pga_2					MH570668	MH570833		MH571100
Lepidocyrtus lanuginosus	LElanuginosus_Pga_3					MH570669	MH570834		MH571101
Lepidocyrtus lanuginosus	LElanuginosus_Pga_5					MH570671	MH570835	MH570969	MH571102
Lepidocyrtus lanuginosus	LElanuginosus_Rad_1	Radom	Grassland-arable land	51°22'49.0"N, 21°06'57.5"E	22/04/2016	MH570672	MH570836	MH570970	MH571103
Lepidocyrtus lanuginosus	LElanuginosus_Rad_2					MH570673	MH570837	MH570971	MH571104
Lepidocyrtus lanuginosus	LElanuginosus_Rad_3					MH570674	MH570838	MH570972	MH571105
Lepidocyrtus lanuginosus	LElanuginosus_Rad_4					MH570675	MH570839	MH570973	MH571106
Lepidocyrtus lanuginosus	LElanuginosus_Rad_5					MH570676	MH570840	MH570974	MH571107
Lepidocyrtus lanuginosus	LElanuginosus_Utr_2	Utrecht	Grassland surround by forest	52°08'50.6"N, 5°12'21.9"E	19/04/2016	MH570677	MH570841		MH571108
Lepidocyrtus lanuginosus	LElanuginosus_Utr_3					MH570678	MH570842		MH571109
Lepidocyrtus lanuginosus	LElanuginosus_Utr_5					MH570679	MH570843	MH570975	MH571110
Lepidocyrtus lanuginosus	LElanuginosus_Vic_1	Vichy	Grassland next to forest	46° 6'24.04"N, 3°22'16.84"E	13/05/2015	MH570680	MH570844	MH570976	
Lepidocyrtus lanuginosus	LElanuginosus_Vic_2					MH570681	MH570845		
Lepidocyrtus lanuginosus	LElanuginosus_Vic_3					MH570682	MH570846	MH570977	MH571111
Lepidocyrtus lanuginosus	LElanuginosus_Vic_4					MH570683	MH570847	MH570978	MH571112
Lepidocyrtus lanuginosus	LElanuginosus_Vic_5					MH570684	MH570848	MH570979	MH571113
Lepidocyrtus lanuginosus	LElanuginosus_Vil_1	near EDAR east (Viladecavalls)	Forest	41°32'49.2"N, 1°57'36.4"E	25/10/2006	MH570685	MH570849		MH571114
Lepidocyrtus lanuginosus	LElanuginosus_Vil_2					MH570686	MH570850		MH571115
Lepidocyrtus lanuginosus	LElanuginosus_Vil_3					MH570687	MH570851		MH571116
Lepidocyrtus lanuginosus	LElanuginosus_Vil_4					MH570688	MH570852	MH570980	MH571117
Lepidocyrtus lanuginosus	LElanuginosus_Vil_5					MH570689	MH570853	MH570981	
Lepidocyrtus lanuginosus	LElanuginosus_Waaa_1	Waake arable land		51°33'47.1"N, 10°03'30.4"E	19/06/2014	MH160136	MH144001	MH143931	MH178000
Lepidocyrtus lanuginosus	LElanuginosus_Waaa_2					MH160137	MH144002		

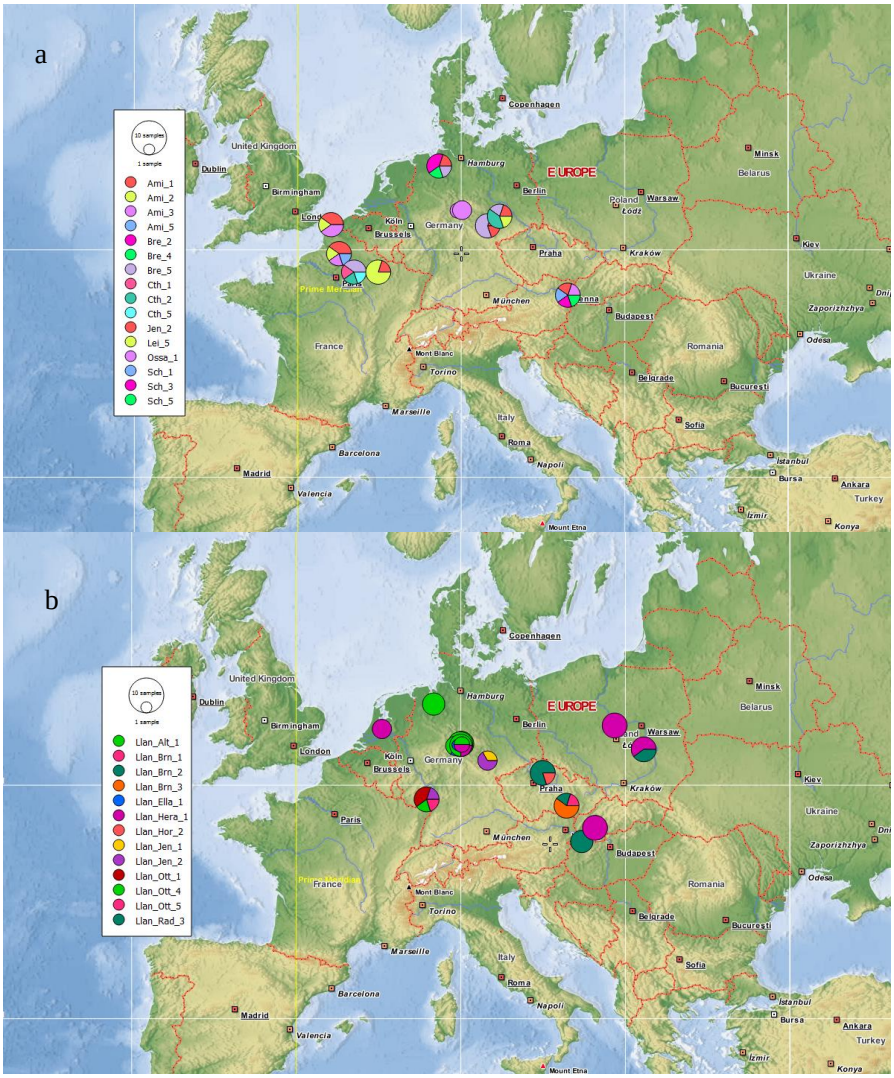
Species	Individual name	Locality	Habitat	Position	Collecting date	28SD1_2	COII	COI	EF1-α
Lepidocyrtus lanuginosus	LElanuginosus_Waaa_3		Arable land and grassland			MH160138	MH144003		
Lepidocyrtus lanuginosus	LElanuginosus_Waaa_4					MH160139	MH144004		
Lepidocyrtus lanuginosus	LElanuginosus_Waaa_5					MH160140	MH144005	MH143932	MH178001
Lepidocyrtus lanuginosus	LElanuginosus_Waaf_1	Waake forest	Forest	51°33'34.0"N, 10°04'14.9"E	19/06/2014	MH160141	MH144006	MH143933	MH178006
Lepidocyrtus lanuginosus	LElanuginosus_Waaf_2					MH160142	MH144007		
Lepidocyrtus lanuginosus	LElanuginosus_Waaf_3					MH160143	MH144008		
Lepidocyrtus lanuginosus	LElanuginosus_Waaf_4					MH160144	MH144009	MH143934	MH178007
Lepidocyrtus lanuginosus	LElanuginosus_Waaf_5					MH160145	MH144010		
Lepidocyrtus paradoxus	LEparadoxus_Mas_1	Masetti	Grassland surround by forest	46° 3'23.22"N, 11°15'29.51"E	04/05/2015	MH160172	MH144037	MH143952	MH178032
Lepidocyrtus cf. *violaceus*	LEviolaceus_Pga_1	Park Gardasee	Grassland surround by forest	45°42'15.73"N, 10°58'23.94"E	04/05/2015	MH160175	MH144040	MH143955	MH178023

Supplementary Table S4.2 Results of the AMOVA on the variation in nucleotide sequences of COII among populations of lineages 3 and 13 (L3 and L13) of the European *Lepidocyrtus lanuginosus* species group. Results were not provided for populations with less than two individuals. Locations refer to Tables 4.1 and Supplementary Table S4.1.

Lineage	Location	Population	N	ss	Eta	h	Hd	Hd (SD)	π	π (SD)	k	D*	F*	*Fs*	Tajima's D
L3	Ami	CE_1	5	16	16	4	0.9	0.161	0.0138	0.00332	9.4	1.64189	1.75174	1.502	1.64189
	Bre	CE_3	5	18	18	4	0.9	0.161	0.01351	0.00318	9.2	0.47737	0.51048	1.467	0.47737
	Cth	CE_4	5	18	19	4	0.9	0.161	0.01322	0.00348	9	0.06473	0.03465	1.432	-0.0971
	Jen	CE_6	5	15	15	3	0.7	0.218	0.01028	0.00359	7	-0.20309	-0.21638	3.005	-0.20309
	Lei	CE_7	5	18	19	4	0.9	0.161	0.01204	0.00304	8.2	-0.58259	-0.65827	1.281	-0.74443
	Ossa	CE_8	1												
	Ossg	CE_9	2	0	0	1	0	0	0	0	0				
	Oyp	CE_10	5	15	15	3	0.8	0.164	0.01233	0.00314	8.4	1.21852	1.29828	3.405	1.21852
	Sch	CE_11	5	15	15	5	1	0.126	0.0094	0.00371	6.4	-0.81235	-0.86552	-0.964	-0.81235
	Smen	CE_12	5	15	15	2	0.4	0.237	0.00881	0.00523	6	-1.21852	-1.29828	5.342	-1.21852
	Waaa	CE_14	3	0	0	1	0	0	0	0	0				
L13															
	Alt	CE_15	4	0	0	1	0	0	0	0	0				
	Billa	CE_16	3	0	0	1	0	0	0	0	0				
	Billf	CE_17	5	0	0	1	0	0	0	0	0				
	Billg	CE_18	4	0	0	1	0	0	0	0	0				
	Brn	CE_19	5	37	37	3	0.7	0.218	0.02203	0.01265	15	-1.16701	-1.26058	4.781	-1.16701
	Depa	CE_20	5	0	0	1	0	0	0	0	0				
	Ella	CE_22	3	1	1	2	0.667	0.314	0.00098	0.00046	0.667				
	Hera	CE_24	4	24	24	2	0.5	0.265	0.01762	0.00935	12	-0.85786	-0.89474	6.118*	-0.85786
	Hor	CE_27	5	14	14	2	0.4	0.237	0.00822	0.00488	5.6	-1.21472	-1.29226	5.145	-1.21472
	Jen	CE_28	3	2	2	2	0.667	0.314	0.00196	0.00092	1.333				
	Kut	CE_29	5	0	0	1	0	0	0	0	0				
	Lebe	CE_30	4	0	0	1	0	0	0	0	0				

Lineage	Location	Population	N	ss	Eta	h	Hd	Hd (SD)	π	π (SD)	k	D*	F*	*Fs*	Tajima's D
	Niea	CE_31	5	0	0	1	0	0	0	0	0				
	Ossa	CE_32	1												
	Ott	CE_34	5	40	42	4	0.9	0.161	0.03172	0.00661	21.6	0.68692	0.71059	2.905	0.53759
	Rad	CE_35	5	1	1	2	0.6	0.175	0.00088	0.00026	0.6	1.22474	1.15728	0.626	1.22474
	Utr	CE_36	3	0	0	1	0	0	0	0	0				
	Waaa	CE_38	2	24	24	2	1	0.5	0.03524	0.01762	24				

N, number of sequences; h, number of haplotypes; ss, number of segregating sites; Eta, total number of mutations; Hd, haplotype diversity; π, nucleotide diversity; k, average number of pairwise nucleotide differences; D*, Fu and Li's D* test statistic; F*, Fu and Li's F* test statistic; Fs, Fu's Fs statistic; SD, standard deviation.*$0.02 < p < 0.05$.



Supplementary Figure S4.1 Distribution of haplotypes of Lineage 3 (L3, a) and L13 (b) in Central Europe. Pie charts represent the frequency of COII lineages in each location. Circle size is proportional to the number of individuals in each location.

Chapter V

General Discussion

The DNA-based analyses presented in this thesis advances our views on the biodiversity, ecology, phylogeny, and phylogeography of European *L. lanuginosus* species group. Two nuclear and two mitochondrial markers demonstrated high intra- and inter-specific genetic divergences, indicating several cryptic species / lineages in both *L. cyaneus* and *L. lanuginosus* (**Chapter II and III**). These lineages differ in habitat preferences (**Chapter II**) and occur in different geographic regions (**Chapter III and IV**). Phylogenetic analyses demonstrated that the European *L. lanuginosus* species group is monophyletic but the two morphologically described species *L. cyaneus* and *L. lanuginosus* are both paraphyletic, thus questioning the current morphological species hypothesis and body color as species marker in this group (**Chapter III**). This study also showed that Southern (SE) and Southwestern Europe (SWE) regions are rich in endemic lineages which are genetically very distinct from lineages in Central Europe (CE), thus rejecting the hypotheses that *L. lanuginosus* (and Collembola in general) went extinct north of the Alps during Quaternary glaciation and recolonized CE from SE and SWE thereafter (**Chapter IV**). At least two lineages co-occur in large areas of CE and both diverged during the Pleistocene and Holocene, which may be due to Quaternary dispersal and isolation and recent human-mediated colonization (**Chapter IV**).

5.1 Incongruence of morphological and molecular species delimitation

It is generally accepted that species are separately evolving metapopulation lineages (the primary defining property of species) and different lines of data (e.g., intrinsic reproductive isolation, diagnosability, and monophyly) provide evidence for defining species (secondary species criteria) (De Queiroz, 2007). Morphological and genetic analyses provide two lines of evidence for species delimitation, but yielded conflicting results for the European *L. lanuginosus* species group in this thesis. DNA-based methods disagreed with the three species hypothesis based on the morphological species concept (MSC) using mainly body color as diagnostic character (Hopkin, 2007; Mateos, 2012, 2008). Rather, the results suggest the existence of nineteen cryptic species / lineages (L1–19) in the species group (**Chapter III**). Similar to other Collembola species groups, such as *Protaphorura* (Sun et al., 2017) and *Coecobrya* (F. Zhang et al., 2018), a threshold of around 5% intraspecific K2P distances were identified in the *L. lanuginosus* species group, which exceeded the widely used 2–3% distance threshold between species (Hebert et al., 2003). The genetic distances within populations of *L. lanuginosus* lineage 13 (**Chapter III**) reached 5% in COI, but whether gene flow exists in different haplotypes within these populations remains unclear. More variable genetic markers such as microsatellites (Rastorgueff et al., 2016; Spinsanti et al., 2006; van der Wurff et al., 2005) and a large set of informative markers such as single nucleotide polymorphisms (SNPs, Brumfield et al., 2003; Shaffer and Thomson, 2007) are needed to explore the gene flow among these lineages.

The phylogenetic analysis further rejected the monophyly of each species (the phylogenetic species concept, PSC), which strongly questions the taxonomic value of body color in the *L. lanuginosus* species group (**Chapter III**). Unlike the genus *Entomobrya* (Collembola: Entomobryidae) and Puerto Rican *Lepidocyrtus*, (Ding et al., 2018; Soto-Adames, 2002), the European *L. lanuginosus* species group is much less diverse in body color patterns, and other morphological characters such as chaetae (Mateos, 2012) may be more helpful in distinguishing these lineages. Integrating genetic and morphological characters is assumed to greatly help strengthening species hypotheses in Collembola (Katz et al., 2015; Sun et al., 2017; Yu et al., 2016; F. Zhang et al., 2018). However, morphological delimitation of Collembola species heavily relies on taxonomic experts and some important morphological characters are delicate and sometimes lost during specimen preservation and slide preparation. Alternative laboratory hybridization tests (c.f., Skarzyński, 2005, 2004) are helpful to confirm the reproductive isolation of these lineages (the BSC), thus provide one of the most important evidences to support the unified species concept (De Queiroz, 2007).

5.2 Environmental filtering of cryptic species of Collembola

The first study (**Chapter II**) compared, for the first time, genetic divergence among populations of a Collembola species from different but connected habitats. The morphologically defined species *L. lanuginosus* was suggested as habitat generalist (Auclerc et al., 2009; Cicconardi et al., 2010; Ferlian et al., 2015; Heiniger et al., 2015; Querner et al., 2013) and was sampled from arable lands, grasslands and forests in this thesis. However, the DNA barcoding gene COI (Herbert 2003, 2004), as well as COII, 28S D1–2 and EF1-α, clustered the individuals to three distinct lineages, i.e., *L. lanuginosus* L1, L2, and L3. These three lineages were included to the European-wide data set of **Chapter III** for DNA taxonomic analysis and were again assigned to three cryptic species / lineages, here L13, L16, and L15, respectively. The habitat preference analysis indicated that only one of the three lineages, i.e., the *L. lanuginosus* L1 in **Chapter II** (L13 in **Chapter III)**, occurred in all three habitats with a preference for arable land but was also common in grasslands across Central Europe (**Chapter III**, Figure 3.1). By using the BLAST function of National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/genbank/), we detected several COI sequences (e.g., accession numbers TUZ031549–TUZ0355) which are nearly identical with this common lineage from different types of forests (Anslan and Tedersoo, 2015), strongly suggesting L13 of the *L. lanuginosus* species group is a habitat generalist.

Another lineage (*L. lanuginosus* L2 in **Chapter II** and L16 in **Chapter III**) was restricted to forests and was never found in grasslands and arable fields in this study. This suggests that at least one lineage of the *L. lanuginosus* suggesting a restricted ecological niche. Similar to the morphospecies *L. lanuginosus*, the morphospecies *L. cyaneus* also occurred in open habitats (Salmon et al., 2014), e.g., grasslands (Auclerc et al., 2009; Migliorini et al., 2003) and arable field (Scheunemann et al., 2015), but they were also collected from forests (Heidemann et al., 2014; Urbanovičová et al., 2014). Considering the high diversity of cryptic

species / lineages and large genetic distances among these lineages, genetic analysis of specimens of *L. cyaneus* from a wider range of locations are necessary to test whether the populations from different habitats and studies belong to a widespread lineage or to distinct genetic lineages.

The species distribution of Collembola significantly links to their morphological traits and environmental factors (Kováč et al., 2016; Salmon et al., 2014). Some species are restricted to particular habitats while others occupy a broad ecological niche (Auclerc et al., 2009). The three lineages of *L. lanuginosus* share a number of morphological traits, such as body size, shape, scales, length of legs and antennae (Mateos, 2012), but differed in their distribution pattern (**Chapter II**). Thus, rather than morphological traits, the distribution of the three lineages of *L. lanuginosus* were more closely related to environmental characteristics. Further studies including investigating physiological characteristics, food resources, and the resistance to soil humidity and disturbances are needed to further understand ecological differentiation. DNA barcoding has revealed that a widely distributed butterfly species in fact comprised a complex of more than ten food plant specialist species (Hebert et al., 2004), and the generalist flower beetle *Mordellistena convicta* has been shown to comprise six cryptic species with distinct sympatric host ranges (Blair et al., 2005), suggesting that 'generalist' species in fact may comprise cryptic complexes of specialist species. DNA barcoding is more accurate and efficient than morphological identification in delimiting these cryptic lineages, thus ecophysiological experiments involving DNA barcoding and laboratory investigations on their physiological characteristics are needed to understand the effects of environmental factors on the distribution patterns of cryptic species, and this is most important in Collembola.

5.3 Geographic and environmental factors limiting gene flow in Collembola

The monophyly of the two morphologically defined species *L. cyaneus* and *L. lanuginosus* was rejected, rather, lineages assigned to each morphospecies represented independent clades that have the same body color (**Chapter II & III**). The biogeographic analysis of all these lineages further rejected a general, wide-spread distribution pattern across-Europe of the two morphospecies *L. cyaneus* and *L. lanuginosus* (Salmon et al., 2014; https://fauna-eu.org), but showed that distinct lineages were distributed in each of the three sampling regions, i.e., SE, SWE, and CE (**Chapter IV**, Figure 4.1). SE and SWE are rich in genetically endemic lineages, supporting the notion that these regions represent hotspots of Collembola diversity (Deharveng et al., 2008; Fiera and Ulrich, 2012). However, not a single lineage was distributed across all three sampling regions, and lineages from SE and SWE showed small range sizes, in contrast to three lineages from CE that were widely distributed. These results indicate that the Pyrenees and Alps function as dispersal barriers, and reduce gene flow among the three sampling regions as well as within SE and SWE, probably due to the high environmental heterogeneity in SE and SWE than in CE.

Landscape barriers and geographical distances that restrict gene flow result in isolation-by-distance

(IBD; Wright, 1943). Selection can drive phenotypic divergence between populations inhabiting different environments and restrict the role of dispersers moving between them to limit gene flow, resulting in isolation-by-environment (IBE, Wang and Summers, 2010). IBD was supported by the population genetic analysis of the Collembola species *Orchesella cincta* (van der Wurff et al., 2005), but was not supported in this thesis. Lineages in each sampling region were non-monophyletic, with genetically closely related lineages being geographically distant (**Chapter IV**, Figures 4.1 and 4.3); genetically distinct lineages overlapped in range size and even coexisted in one habitat (L3 and L13 in **Chapter IV**). Apparently, the genetic pattern in the *L. lanuginosus* species group is more complex than in *O. cincta.* However, IBE is supported as three lineages of the morphospecies *L. lanuginosus* showed preferences for a certain habitat, despite the close vicinity of habitats (**Chapter II**) and free dispersal of this fast dispersing species (Auclerc et al., 2009; Ponge et al., 2006) re-colonizing habitats quickly after disturbances (González-Macé and Scheu, 2018). Rather, certain lineages of the *L. lanuginosus* species group were either restricted to certain habitats in the same region (**Chapter II**) or to different environmental regions (the three sampling regions in **Chapter III & IV**). Thus in general, the results suggest that IBE plays a more important role than IBD in shaping the distribution of the lineages of the *L. lanuginosus* species group.

Divergence time estimations allow to infer the relevance of historical geography and climatic changes for lineage divergence. The main lineages of the *L. lanuginosus* species group diverged during the late Miocene, indicating that the common ancestors of these lineages colonized Europe more than ten million years ago. Climate changes in the late Miocene, i.e., slow transition from warm and humid to cool and dry climate (Zachos et al., 2001), the increase of seasonality (Bruch et al., 2007) and fire frequency (Beerling and Osborne, 2006), and the decrease of atmospheric CO_2 concentration (Mosbrugger et al., 2005), contributed the expansion of C_4 grasslands in Europe (Kürschner et al., 2008) and significant shifts of forest types (Utescher et al., 2007). Populations of Collembola may have experienced large scale dispersal and later on long-term isolation across Europe, particularly in the Mediterranean region (Cicconardi et al., 2010). Quaternary glaciation was indicated to play a minor role for the survival of Collembola, as lineages of a number of Collembola species from CE separated from those in SE and SWE long before the Quaternary (Cicconardi et al., 2010; Porco et al., 2012; Timmermans et al., 2005; von Saltzwedel et al., 2017, 2016; B. Zhang et al., 2018). However, a number of haplotypes of two widely distributed lineages (L3 and L13 of the *L. lanuginosus* species group) dated back to the Pleistocene and Holocene, suggesting that Quaternary glaciation drove the dispersal and diversification of Collembola within CE (**Chapter IV**). Identical or very similar haplotypes of these two lineages inhabited grasslands hundreds of kilometers apart. This is possibly due to recent human-introduced species invasion as indicated by population genetic analyses on Collembola (Cicconardi et al., 2017; Porco et al., 2013). More variable genetic markers are needed to confirm this hypothesis.

5.4 Future perspectives

Morphological identification of Collembola to species level requires great amount of time, money and taxonomic experts. As revealed by this thesis and previous genetic studies (Cicconardi et al., 2010; Porco et al., 2012), conventional microscopic examination has ignored cryptic species and underestimated their biodiversity. As alternative, Sanger sequencing of hundreds to thousands of specimens extracted from soil or litter samples in ecological studies is time consuming and expensive. Recently DNA metabarcoding has been proposed as an efficient way in delimiting diagnostic operational taxonomic units (OTUs) of certain DNA fragments from bulk Collembola community samples (Saitoh et al., 2016; Taberlet et al., 2012). However, unlike the exact 90% (Saitoh et al., 2016) or 97% sequence similarity threshold (Arribas et al., 2016; Yu et al., 2012), this thesis suggested a range of 90.5-94.5% sequence similarity threshold in COI according to the 5.5–9.5% K2P distances gap identified in the *L. lanuginosus* species group (**Chapter III**). A universal threshold of genetic distance for species delimitation was not recommended in Collembola, rather the barcoding gaps or species boundaries should be based on calculations within specific groups of species (Sun et al., 2017; F. Zhang et al., 2018), particularly in species comprising cryptic species / lineages. Thus, rather than just deliver OTUs under a common threshold, genetic sequences should be assigned to certain morphological taxa before delivering OTUs under specific thresholds in metabarcoding studies. The effects of using certain thresholds for species delimitation on false positive (over-splitting) and false negative (over-lumping) rates should be evaluated in future studies (Meyer and Paulay, 2005). Additionally, DNA metabarcoding can explore within-population haplotype and nucleotide diversity thus benefiting phylogeographic studies of sympatric species / lineages.

Studies on the molecular phylogeny of Collembola are mainly based on one or few selected genes obtained by Sanger sequencing (but see, Carapelli et al., 2014), while next-generation sequencing has been recommend in biological applications due to the advantages of faster and larger data generation and lower sequencing costs (Schuster, 2008). Phylogenetic analyses based on hundreds to thousands of genes obtained by high-throughput sequencing and appropriate analytical methods have produced statistically robust and congruent results resolving phylogenies of arthropods (Misof et al., 2014; Regier et al., 2010; Rota-Stabelli et al., 2013; Thomas et al., 2018). Mitochondrial genomes provided an excellent data source to recover inter- and intra-ordinal phylogeny in insects (reviewed by Cameron, 2014) such as Diptera (Cameron et al., 2007), Coleoptera (Doucet-Beaupré et al., 2010; Song et al., 2010; Timmermans et al., 2010; Timmermans and Vogler, 2012), Neuropterida (Cameron et al., 2009), Hemiptera (Song et al., 2012), and Acari (Xue et al., 2017). However, base compositional heterogeneity and among-site rate variation affect their phylogenetic inference, causing unexpected relationships but with high branch support (Song et al., 2010). Integrating mitochondrial genomes, nuclear gene sequences, and morphology and utilizing refined analytical methods are necessary to understand insect evolution (Cameron, 2014), as well as other arthropods including Collembola.

Many species or genera of Collembola, as the morphospecies *L. cyaneus* and *L. lanuginosus* in this thesis, are globally distributed, even though they are assumed to have limited dispersal capacity (Hopkin, 1997). The long-term historical geographic and climate changes and dispersal of species by humans both shaped the biogeographical distribution pattern of Collembola which very likely differs from that of plants and vertebrates. Population genetic analyses based on genomic data and global scale sampling of a range of taxa is needed to deepen understanding of dispersal, speciation, and evolution of Collembola. Such studies will reveal if evolutionary patterns in Collembola resemble those of above-ground species, or show specific patterns typical for below-ground species.

Conclusions

Genetic analyses based on sampling across-Europe allowed novel insight into the diversity, ecology, phylogeny, and phylogeography of species of the European *L. lanuginosus* species group. The genetic diversity and structure of two species of the *L. lanuginosus* species group, i.e. *L. cyaneus* and *L. lanuginosus*, correlated with habitat types, geographical regions, and historical climate changes. DNA-based approaches revealed distinct cryptic species / lineages in each of these morphologically defined species, and demonstrated their paraphyly, but supported monophyly of the *L. lanuginosus* species group. Body color, which has been considered as one of the most important characters in species delimitation in the *L. lanuginosus* species group, was proven to be an improper species and lineage marker. COI, COII, and EF1-αseparated main lineages of the *L. lanuginosus* species group, but 28S is too conserved and neither allows reconstructing phylogeny nor delimiting these closely related lineages. This thesis suggests that historical geographic and climatic changes dating back to the late Miocene, but also recent human-mediated dispersal shaped the present-day distribution of Collembola species. The isolation of lineages of the *L. lanuginosus* species group was not related to geographic distances. In contrast, environmental differences, among and within sampling regions, was assumed to be the main factors impeding gene flow. Climate changes in the Miocene and the accompanied expansion of grasslands and shifts of forest types may have been responsible for the dispersal and later isolation of ancestral lineages of the *L. lanuginosus* species group. Quaternary glaciation presumably contributed little to the sorting of lineages of the *L. lanuginosus* species group and to the dispersal from CE to SE and SWE or vise visa, but likely shaped the distribution of Collembola within CE. Further, the results suggest that human-mediated dispersal contributed to the colonization of distant localities by similar or identical haplotypes. Overall, the results of this thesis suggest that soil animals likely experienced different and more complex evolutionary forces than aboveground animals and plants shaping their genetic diversification and biogeographic distribution. Global scale

phylogeographic studies on a number of Collembola species are needed for a deeper understanding of the dispersal, speciation, and evolution of Collembola and soil animals in general.

References

Anslan, S., Tedersoo, L., 2015. Performance of cytochrome c oxidase subunit I (COI), ribosomal DNA Large Subunit (LSU) and Internal Transcribed Spacer 2 (ITS2) in DNA barcoding of Collembola. Eur. J. Soil Biol. 69, 1–7. doi:10.1016/j.ejsobi.2015.04.001

Arribas, P., Andújar, C., Hopkins, K., Shepherd, M., Vogler, A.P., 2016. Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil. Methods Ecol. Evol. 7, 1071–1081. doi:10.1111/2041-210X.12557

Auclerc, A., Ponge, J.F., Barot, S., Dubs, F., 2009. Experimental assessment of habitat preference and dispersal ability of soil springtails. Soil Biol. Biochem. 41, 1596–1604. doi:10.1016/j.soilbio.2009.04.017

Beerling, D.J., Osborne, C.P., 2006. The origin of the savanna biome. Glob. Chang. Biol. 12, 2023–2031. doi:10.1111/j.1365-2486.2006.01239.x

Blair, C.P., Abrahamson, W.G., Jackman, J.A., Tyrrell, L., 2005. Cryptic speciation and host-race formation in a purportedly generalist tumbling flower beetle. Evolution (N. Y). 59, 304–316. doi:10.1111/j.0014-3820.2005.tb00991.x

Bruch, A.A., Uhl, D., Mosbrugger, V., 2007. Miocene climate in Europe — Patterns and evolution. Palaeogeogr. Palaeoclimatol. Palaeoecol. 253, 1–7. doi:10.1016/j.palaeo.2007.03.030

Brumfield, R.T., Beerli, P., Nickerson, D.A., Edwards, S. V., 2003. The utility of single nucleotide polymorphisms in inferences of population history. Trends Ecol. Evol. 18, 249–256. doi:10.1016/S0169-5347(03)00018-1

Cameron, S.L., 2014. Insect mitochondrial genomics: implications for evolution and phylogeny. Annu. Rev. Entomol. 59, 95–117. doi:10.1146/annurev-ento-011613-162007

Cameron, S.L., Lambkin, C.L., Barker, S.C., Whiting, M.F., 2007. A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. Syst. Entomol. 32, 40–59. doi:10.1111/j.1365-3113.2006.00355.x

Cameron, S.L., Sullivan, J., Song, H., Miller, K.B., Whiting, M.F., 2009. A mitochondrial genome phylogeny of the Neuropterida (lace-wings, alderflies and snakeflies) and their relationship to the other holometabolous insect orders. Zool. Scr. 38, 575–590. doi:10.1111/j.1463-6409.2009.00392.x

Carapelli, A., Convey, P., Nardi, F., Frati, F., 2014. The mitochondrial genome of the antarctic springtail *Folsomotoma octooculata* (Hexapoda; Collembola), and an update on the phylogeny of collembolan lineages based on mitogenomic data. Entomologia 2. doi:10.4081/entomologia.2014.190

Cicconardi, F., Borges, P.A. V., Strasberg, D., Oromí, P., López, H., Pérez-Delgado, A.J., Casquet, J., Caujapé-Castells, J., Fernández-Palacios, J.M., Thébaud, C., Emerson, B.C., 2017. MtDNA metagenomics reveals large-scale invasion of belowground arthropod communities by introduced species. Mol. Ecol. 26, 3104–3115. doi:10.1111/mec.14037

Cicconardi, F., Nardi, F., Emerson, B.C., Frati, F., Fanciulli, P.P., 2010. Deep phylogeographic divisions and long-

term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. Mol. Ecol. 19, 386–400. doi:10.1111/j.1365-294X.2009.04457.x

De Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886. doi:10.1080/10635150701701083

Deharveng, L., D'Haese, C.A., Bedos, A., 2008. Global diversity of springtails (Collembola; Hexapoda) in freshwater. Hydrobiologia 595, 329–338. doi:10.1007/s10750-007-9116-z

Ding, Y., Yu, D., Guo, W., Li, J., Zhang, F., 2018. Molecular phylogeny of *Entomobrya* (Collembola: Entomobryidae) from China: Color pattern groups and multiple origins. Insect Sci. 1–11. doi:10.1111/1744-7917.12559

Doucet-Beaupré, H., Breton, S., Chapman, E.G., Blier, P.U., Bogan, A.E., Stewart, D.T., Hoeh, W.R., 2010. Mitochondrial phylogenomics of the Bivalvia (Mollusca): searching for the origin and mitogenomic correlates of doubly uniparental inheritance of mtDNA. BMC Evol. Biol. 10, 50. doi:10.1186/1471-2148-10-50

Ferlian, O., Klarner, B., Langeneckert, A.E., Scheu, S., 2015. Trophic niche differentiation and utilisation of food resources in collembolans based on complementary analyses of fatty acids and stable isotopes. Soil Biol. Biochem. 82, 28–35. doi:10.1016/j.soilbio.2014.12.012

Fiera, C., Ulrich, W., 2012. Spatial patterns in the distribution of European springtails (Hexapoda: Collembola). Biol. J. Linn. Soc. 105, 498–506. doi:10.1111/j.1095-8312.2011.01816.x

González-Macé, O., Scheu, S., 2018. Response of Collembola and Acari communities to summer flooding in a grassland plant diversity experiment. PLoS One 13, e0202862. doi:10.1371/journal.pone.0202862

Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B Biol. Sci. 270, 313–321. doi:10.1098/rspb.2002.2218

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., Hallwachs, W., 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Natl. Acad. Sci. 101, 14812–14817. doi:10.1073/pnas.0406166101

Heidemann, K., Hennies, A., Schakowske, J., Blumenberg, L., Ruess, L., Scheu, S., Maraun, M., 2014. Free-living nematodes as prey for higher trophic levels of forest soil food webs. Oikos 123, 1199–1211. doi:10.1111/j.1600-0706.2013.00872.x

Heiniger, C., Barot, S., Ponge, J.-F., Salmon, S., Meriguet, J., Carmignac, D., Suillerot, M., Dubs, F., 2015. Collembolan preferences for soil and microclimate in forest and pasture communities. Soil Biol. Biochem. 86, 181–192. doi:10.1016/j.soilbio.2015.04.003

Hopkin, S.P., 2007. A Key to the Springtails (Collembola) of Britain and Ireland, Field Studies Council (AIDGAP Project). FSC Publicatio, Shrewsbury.

Hopkin, S.P., 1997. Biology of the springtails: (Insecta: Collembola). Oxford University Press, Oxford, UK.

Katz, A.D., Giordano, R., Soto-Adames, F.N., 2015. Operational criteria for cryptic species delimitation when evidence is limited, as exemplified by North American *Entomobrya* (Collembola: Entomobryidae). Zool. J. Linn. Soc. 173, 818–840. doi:10.1111/zoj.12220

Kováč, Ľ., Parimuchová, A., Miklisová, D., 2016. Distributional patterns of cave Collembola (Hexapoda) in association with habitat conditions, geography and subterranean refugia in the Western Carpathians. Biol.

J. Linn. Soc. 119, 571–592. doi:10.1111/bij.12555

Kürschner, W.M., Kvaček, Z., Dilcher, D.L., 2008. The impact of Miocene atmospheric carbon dioxide fluctuations on climate and the evolution of terrestrial ecosystems. Proc. Natl. Acad. Sci. 105, 449–453. doi:10.1073/pnas.0708588105

Mateos, E., 2012. The European *Lepidocyrtus lanuginosus* group (Collembola: Entomobryidae), definition and description of a new species from Spain. Zootaxa 3570, 69–81.

Mateos, E., 2008. The European *Lepidocyrtus* Bourlet, 1839 (Collembola: Entomobryidae). Zootaxa 1769, 35–59.

Meyer, C.P., Paulay, G., 2005. DNA barcoding: Error rates based on comprehensive sampling. PLoS Biol. 3, 1–10. doi:10.1371/journal.pbio.0030422

Migliorini, M., Fanciulli, P.P., Bernini, F., 2003. Comparative analysis of two edaphic zoocoenoses (Acari Oribatida; Hexapoda Collembola) in the area of Orio al Serio Airport (Bergamo, northern Italy). Pedobiologia. 47, 9–18. doi:10.1078/0031-4056-00164

Misof, B., Liu, S., Meusemann, K., Peters, R.S., et al., 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science (80-.). 346, 763–767. doi:10.1126/science.1257570

Mosbrugger, V., Utescher, T., Dilcher, D.L., 2005. Cenozoic continental climatic evolution of Central Europe. Proc. Natl. Acad. Sci. 102, 14964–14969. doi:10.1073/pnas.0505267102

Ponge, J.F., Dubs, F., Gillet, S., Sousa, J.P., Lavelle, P., 2006. Decreased biodiversity in soil springtail communities: the importance of dispersal and landuse history in heterogeneous landscapes. Soil Biol. Biochem. 38, 1158–1161. doi:10.1016/j.soilbio.2005.09.004

Porco, D., Bedos, A., Greenslade, P., Janion, C., Skarzyński, D., Stevens, M.I., Jansen Van Vuuren, B., Deharveng, L., 2012. Challenging species delimitation in Collembola: cryptic diversity among common springtails unveiled by DNA barcoding. Invertebr. Syst. 26, 470–477. doi:10.1071/IS12026

Porco, D., Decaëns, T., Deharveng, L., James, S.W., Skarzyński, D., Erséus, C., Butt, K.R., Richard, B., Hebert, P.D.N., 2013. Biological invasions in soil: DNA barcoding as a monitoring tool in a multiple taxa survey targeting European earthworms and springtails in North America. Biol. Invasions 15, 899–910. doi:10.1007/s10530-012-0338-2

Querner, P., Bruckner, A., Drapela, T., Moser, D., Zaller, J.G., Frank, T., 2013. Landscape and site effects on Collembola diversity and abundance in winter oilseed rape fields in eastern Austria. Agric. Ecosyst. Environ. 164, 145–154. doi:10.1016/j.agee.2012.09.016

Rastorgueff, P.-A., Arnal, V., Montgelard, C., Monsanto, D.M., Groenewald, C.W., Haddad, W.A., Dubois, M.-P., Jansen van Vuuren, B., 2016. Characterization of 21 polymorphic microsatellite loci for the collembolan *Cryptopygus antarcticus* travei from the sub-Antarctic Prince Edward Islands. Biochem. Syst. Ecol. 64, 136–141. doi:10.1016/j.bse.2015.12.003

Regier, J.C., Shultz, J.W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J.W., Cunningham, C.W., 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature 463, 1079–1083. doi:10.1038/nature08742

Rota-Stabelli, O., Daley, A.C., Pisani, D., 2013. Molecular timetrees reveal a cambrian colonization of land and a new scenario for ecdysozoan evolution. Curr. Biol. 23, 392–398. doi:10.1016/j.cub.2013.01.026

Saitoh, S., Aoyama, H., Fujii, S., Sunagawa, H., Nagahama, H., Akutsu, M., Shinzato, N., Kaneko, N., Nakamori, T., 2016. A quantitative protocol for DNA metabarcoding of springtails (Collembola). Genome 59, 705–723. doi:10.1139/gen-2015-0228

Salmon, S., Ponge, J.F., Gachet, S., Deharveng, L., Lefebvre, N., Delabrosse, F., 2014. Linking species, traits and habitat characteristics of Collembola at European scale. Soil Biol. Biochem. 75, 73–85. doi:10.1016/j.soilbio.2014.04.002

Scheunemann, N., Digel, C., Scheu, S., Butenschoen, O., 2015. Roots rather than shoot residues drive soil arthropod communities of arable fields. Oecologia 179, 1135–1145. doi:10.1007/s00442-015-3415-2

Schuster, S.C., 2008. Next-generation sequencing transforms today's biology. Nat. Methods 5, 16–18. doi:10.1038/nmeth1156

Shaffer, H.B., Thomson, R.C., 2007. Delimiting species in recent radiations. Syst. Biol. 56, 896–906. doi:10.1080/10635150701772563

Skarzyński, D., 2005. Taxonomy of the ceratophysellan lineage (Collembola: Hypogastruridae) in the light of laboratory hybridization studies. Zootaxa 1043, 47–59. doi:10.11646/zootaxa.1043.1.4

Skarzyński, D., 2004. Taxonomic status of *Schaefferia emucronata* Absolon, 1900 and *Schaefferia willemi* (Bonet. 1930)(Collembola, Hypogastruridae) in the light of hybridisation studies. Dtsch. Entomol. Zeitschrift 51, 81–86.

Song, H., Sheffield, N.C., Cameron, S.L., Miller, K.B., Whiting, M.F., 2010. When phylogenetic assumptions are violated: base compositional heterogeneity and among-site rate variation in beetle mitochondrial phylogenomics. Syst. Entomol. 35, 429–448. doi:10.1111/j.1365-3113.2009.00517.x

Song, N., Liang, A.-P., Bu, C.-P., 2012. A molecular phylogeny of Hemiptera inferred from mitochondrial genome sequences. PLoS One 7, e48778. doi:10.1371/journal.pone.0048778

Soto-Adames, F.N., 2002. Molecular phylogeny of the Puerto Rican *Lepidocyrtus* and *Pseudosinella* (Hexapoda: Collembola), a validation of Yoshii's "color pattern species." Mol. Phylogenet. Evol. 25, 27–42. doi:10.1016/S1055-7903(02)00250-6

Spinsanti, G., Nardi, F., Frati, F., 2006. Isolation of novel microsatellite loci in *Orchesella villosa* (Arthropoda, Collembola). Pedobiologia. 50, 105–109. doi:10.1016/j.pedobi.2005.10.003

Sun, X., Zhang, F., Ding, Y., Davies, T.W., Li, Y., Wu, D., 2017. Delimiting species of *Protaphorura* (Collembola: Onychiuridae): integrative evidence based on morphology, DNA sequences and geography. Sci. Rep. 7, 8261. doi:10.1038/s41598-017-08381-4

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., Willerslev, E., 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol. Ecol. 21, 2045–2050. doi:10.1111/j.1365-294X.2012.05470.x

Thomas, G.W.C., Dohmen, E., Hughes, D.S.T., et al., 2018. The genomic basis of Arthropod diversity. bioRxiv 382945. doi:10.1101/382945

Timmermans, M.J.T.N., Dodsworth, S., Culverwell, C.L., Bocak, L., Ahrens, D., Littlewood, D.T.J., Pons, J., Vogler, A.P., 2010. Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. Nucleic Acids Res. 38. doi:10.1093/nar/gkq807

Timmermans, M.J.T.N., Ellers, J., Mariën, J., Verhoef, S.C., Ferwerda, E.B., Van Straalen, N.M., 2005. Genetic

structure in *Orchesella cincta* (Collembola): Strong subdivision of European populations inferred from mtDNA and AFLP markers. Mol. Ecol. 14, 2017–2024. doi:10.1111/j.1365-294X.2005.02548.x

Timmermans, M.J.T.N., Vogler, A.P., 2012. Phylogenetically informative rearrangements in mitochondrial genomes of Coleoptera, and monophyly of aquatic elateriform beetles (Dryopoidea). Mol. Phylogenet. Evol. 63, 299–304. doi:10.1016/j.ympev.2011.12.021

Urbanovičová, V., Miklisová, D., Kováč, Ľ., 2014. Forest disturbance enhanced the activity of epedaphic Collembola in windthrown stands of the High Tatra mountains. J. Mt. Sci. 11, 449–463. doi:10.1007/s11629-013-2736-z

Utescher, T., Erdei, B., François, L., Mosbrugger, V., 2007. Tree diversity in the Miocene forests of Western Eurasia. Palaeogeogr. Palaeoclimatol. Palaeoecol. 253, 226–250. doi:10.1016/j.palaeo.2007.03.041

van der Wurff, A.W.G., Gols, R., Ernsting, G., van Straalen, N.M., 2005. Population genetic structure of *Orchesella cincta* (Collembola; Hexapoda) in NW Europe, as revealed by microsatellite markers. Pedobiologia. 49, 167–174. doi:10.1016/j.pedobi.2004.10.004

von Saltzwedel, H., Scheu, S., Schaefer, I., 2017. Genetic structure and distribution of *Parisotoma notabilis* (Collembola) in Europe: Cryptic diversity, split of lineages and colonization patterns. PLoS One 12, e0170909. doi:10.1371/journal.pone.0170909

von Saltzwedel, H., Scheu, S., Schaefer, I., 2016. Founder events and pre-glacial divergences shape the genetic structure of European Collembola species. BMC Evol. Biol. 16, 148. doi:10.1186/s12862-016-0719-8

Wang, I.J., Summers, K., 2010. Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. Mol. Ecol. 19, 447–458. doi:10.1111/j.1365-294X.2009.04465.x

Wright, S., 1943. Isolation by Distance. Genetics 28, 114–138.

Xue, X.-F., Dong, Y., Deng, W., Hong, X.-Y., Shao, R., 2017. The phylogenetic position of eriophyoid mites (superfamily Eriophyoidea) in Acariformes inferred from the sequences of mitochondrial genomes and nuclear small subunit (18S) rRNA gene. Mol. Phylogenet. Evol. 109, 271–282. doi:10.1016/j.ympev.2017.01.009

Yu, D., Zhang, F., Stevens, M.I., Yan, Q., Liu, M., Hu, F., 2016. New insight into the systematics of Tomoceridae (Hexapoda, Collembola) by integrating molecular and morphological evidence. Zool. Scr. 45, 286–299. doi:10.1111/zsc.12149

Yu, D.W., Ji, Y., Emerson, B.C., Wang, X., Ye, C., Yang, C., Ding, Z., 2012. Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods Ecol. Evol. 3, 613–623. doi:10.1111/j.2041-210X.2012.00198.x

Zachos, J., Pagani, H., Sloan, L., Thomas, E., Billups, K., 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. Science (80-.). 292, 686–693. doi:10.1126/science.1059412

Zhang, B., Chen, T.W., Mateos, E., Scheu, S., Schaefer, I., 2018. Cryptic species in *Lepidocyrtus lanuginosus* (Collembola: Entomobryidae) are sorted by habitat type. Pedobiologia. 68, 12–19. doi:10.1016/j.pedobi.2018.03.001

Zhang, F., Jantarit, S., Nilsai, A., Stevens, M.I., Ding, Y., Satasook, C., 2018. Species delimitation in the

morphologically conserved *Coecobrya* (Collembola: Entomobryidae): A case study integrating morphology and molecular traits to advance current taxonomy. Zool. Scr. 47, 342–356. doi:10.1111/zsc.12279

Acknowledgements

I owe many thanks to Stefan Scheu for accepting me as a PhD student in his working group and providing such a stimulating work atmosphere. I thank him for many inspiring and motivating discussions of my thesis. I am also very grateful to him for wonderful excursions to Giglio and the Alps which were unforgettable experiences.

I thank Christoph Bleidorn for his efforts as referee for this dissertation. I appreciate Elvira Hörandl, Klaus Hövemeyer, Mark Maraun, and Marko Rohlfs for their willingness and the time they have invested in being part of the committee.

I thank the China Scholarship Council for financing me as a doctoral student in Germany, and the people who helped me with the scholarship application: my master's supervisor Dong-Hui Wu and members of his working group, Liang Chang, Xin Sun, Jun-Ming Liu, and Mei-Xiang Gao.

I would like to give my special thanks to Dr. Eduardo Mateos from the University of Barcelona. He gave me professional introductions in Collembola taxonomy and kind help in taking samples in Spain. I enjoyed a great time in Barcelona, and the skills he taught me greatly improved my work on specimen identification.

I thank Andrea Lambertz and the technicians in the working group, Susanne Böning-Klein, Christel Fischer, Guido Humpert, Dieter Nünchert, Elisabeth Opielka, and Theodora Volovei. Their efficient organization kept my lab work running smoothly.

I thank the colleagues who helped me with field work, Ting-Wen Chen, Bernhard Eitzinger, Odette Gonzalez, Kristin Haynert, Bastian Heimburger, Andreas Klein, and Laura Sanchez. I also want to thank Stefan, Ting-Wen, Ina Schaefer, and Feng Zhang for the inspiring discussions and comments which improved the quality of the thesis very much.

Many thanks to the core members in the paper writing group, Odette, Tamara Hartke, Nico Radermacher, and Garvin Schulz; gathering together really speeded up and improved my writing.

I would like to thank the whole AG Scheu for such a nice working atmosphere. I have really enjoyed my time here. Also thank you my Chinese colleagues, Liang, Jo-Fan Chao, Ting-Wen, Xin Gong, Zhi-Peng Li, Jing-Zhong Lu, Di-Qiu Sun, Xin S., Si-Min Wang, Feng, Hui-Ning Zhang, Shi-Xing Zhou; they have helped me a lot in work and life in Göttingen.

I also want to thank all my Chinese friends in Göttingen, particularly the team members of the Chinese badminton and football clubs; they have made my life in Germany not lonely and boring but colorful and wonderful.

Last, I give my special thanks to my father Xue-Zhen Zhang, my mother Qi-Lian Liao, and younger sister Yan Zhang for their patience, company and love in all situations.

List of publications

Published paper related to this thesis

Bing Zhang, Ting-Wen Chen, Eduardo Mateos, Stefan Scheu, and Ina Schaefer. Cryptic species in *Lepidocyrtus lanuginosus* (Collembola: Entomobryidae) are sorted by habitat type. *Pedobiologia*, 2018, 68: 12-19. Doi: 10.1016/j.pedobi.2018.03.001

Unpublished papers related to this thesis

Bing Zhang, Ting-Wen Chen, Eduardo Mateos, Stefan Scheu, and Ina Schaefer. DNA-based approaches uncover cryptic diversity in the European *Lepidocyrtus lanuginosus* species group (Collembola: Entomobryidae). (Submitted to *Invertebrate Systematics*, minor revision)

Bing Zhang, Ting-Wen Chen, Eduardo Mateos, Stefan Scheu, and Ina Schaefer. Late Miocene-Pliocene diversification and Pleistocene-Holocene colonization shape the phylogeography of the European *Lepidocyrtus lanuginosus* species group (Collembola: Entomobryidae). (In preparation)

Other published papers

Bing Zhang, Liang Chang, Zhen Ni, Xin Sun, and Donghui Wu. Directional migration of three *Desoria* species (Collembola: Isotomidae) on the snow surface in late winter. *European Journal of Soil Biology*, 2017, 81: 64-68. Doi: 10.1016/j.ejsobi.2017.06.009

Bing Zhang, Liang Chang, Zhen Ni, Mac A Callaham Jr., Xin Sun, and Donghui Wu. Effects of land use changes on winter-active Collembola in Sanjiang Plain of China. *Applied Soil Ecology*, 2014, 83, 51-58. Doi: 10.1016/j.apsoil.2014.03.008

Bing Zhang, Zhen Ni, Liang Chang, Haitao, Wu, Xin Sun, and Donghui Wu. A review of the snow-living Collembola. *Acta Ecologica Sinica*, 2014, 34 (8): 1922-1936. (In Chinese with English abstract).

Xin Sun, **Bing Zhang**, Wu Donghui. Two new species and a new country record of *Protaphorura Absolon*, 1901 (Collembola: Onychiuridae) from Northeast China. *Zootaxa*. 2013, 3693 (2): 207-220. Doi: 0.11646/zootaxa.3693.2.7

Michael T. Marx, Xiumin Yan, Xuefeng Wang, Lihong Song, Kehong Wang, **Bing Zhang**, and Donghui Wu. Soil fauna abundance, feeding and decomposition in different reclaimed and natural sites in the Sanjiang Plain wetland, Northeast China. *Wetlands*, 2016, 36 (3): 445-455. Doi: 10.1007/s13157-016-0753-8

Thesis declarations

Declaration of the author's own contribution to manuscripts with multiple authors

Chapter II comprises a manuscript that has been published in a peer-reviewed journal. Samples for **Chapter II** were collected by Ting-Wen Chen, molecular sequence data in **Chapter II** was obtained with the help of Jo-Fan Chao. Ting-Wen Chen and I have equally contributed to the manuscript presented in **Chapter II**. Ting-Wen Chen developed the study, we both collected the data, and I analyzed the data and wrote the manuscript. **Chapter III** is currently submitted to a peer-reviewed journal; I have collected all data.

I am the first author of all manuscripts; I have analyzed the data, written the manuscripts, created tables, figures and supplementary materials and contributed significantly to the study design. All persons contributing to the manuscripts have been named so.

Plagiarism declaration

I, Bing Zhang, declare that I have written this doctoral thesis independently. All persons contributing to the manuscripts have been named so. All sentences or passages quoted from other people's work have been specifically acknowledged by clear cross-referencing.

I have not submitted this thesis in any form for another degree at any university or institution.

_____

Bing Zhang

Göttingen, November 2018

Curriculum vitae

Personal details

Name	Bing Zhang
Birth	June 20, 1989
Nationality	People's Republic of China

Education

10.2014–12.2018 Ph.D. thesis, major in Biology, Georg-August University Goettingen, Germany.
Title: Habitat selection, cryptic diversity, phylogeny, and phylogeography of the European *Lepidocyrtus lanuginosus* species group (Collembola: Entomobryidae).

09.2011–07.2014 Master degree of Management, major in Land Resource Management, College of Earth Sciences, Jilin University, China.
Title: Effects of land use types on biodiversity of winter active Collembola.

09.2007–07.2011 Bachelor degree of Agriculture, major in Agricultural Resources and Environment, College of Plant Sciences, Jilin University, China.
Title: Cloning and genetic transformation of *ZmCIPK*4 and *ZmCIPK*8 in maize.

Teaching

09.2017–11.2018 Supervision of a MSc thesis (Simin Wang)
Georg-August University Goettingen, Germany

12.2017–01.2018 Soil Animal Ecology
Georg-August University Goettingen, Germany

Conferences

04.2018 The 14th Central European Workshop on Soil Zoology, České Budějovice, Czech Republic
Talk: Genetic structure among cryptic species in the European *Lepidocyrtus lanuginosus* group (Collembola: Entomobryidae)

10.2017 Second Global Soil Biodiversity Conference, Nanjing, China
Poster: Cryptic Species or Bad Characters

08.2016 XVII International Colloquium on Soil Zoology and XIV International Colloquium on Apterygota, Nara, Japan
Talk: Genetic variation in European *Lepidocyrtus cyaneus* (Collembola: Entomobryidae) as indicated by nuclear and mitochondrial markers

Internships

09.2015 Taxonomy of *Lepidocyrtus*, University of Barcelona, Spain.

06.2012–05.2014 Visiting student, Soil Biodiversity Group, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China