The sibling species *Leptidea juvernica* and *L. sinapis* (Lepidoptera, Pieridae) in the Balkan Peninsula: ecology, genetic structure, and morphological variation

Nikolay Shtinkov[a, *], Zdravko Kolev[b], Roger Vila[c], Vlad Dincă[c, d]

[a] Department of Physics, University of Ottawa, 150 Louis Pasteur, Ottawa, ON, K1N 6N5, Canada. E-mail address: nshtinkov@uottawa.ca
[b] National Museum of Natural History, 1 Tsar Osvoboditel Blvd, 1000, Sofia, Bulgaria. E-mail address: zkolev@nmnhs.com
[c] Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37, 08003, Barcelona, Spain. E-mail address: roger.vila@csic.es
[d] Biodiversity Institute of Ontario, University of Guelph, Guelph, ON, N1G 2W1, Canada. E-mail address: vdinca@uoguelph.ca

[*] Corresponding author. Tel.: +1-613-562-5800 ext. 6547; e-mail address: nshtinkov@uottawa.ca

Abstract

The Wood White butterfly *Leptidea sinapis* and its more recently discovered sibling species *L. reali* and *L. juvernica* have emerged as a model system for studying the speciation and evolution of cryptic species, as well as their ecological interactions in conditions of sympatry. *Leptidea sinapis* is widely distributed from Western Europe to Central Asia while the synmorphic *L. juvernica* and *L. reali* have allopatric distributions, both occurring in sympatry with *L. sinapis* and exhibiting an intricate, regionally variable ecological niche separation. Until now, the Balkan Peninsula remained one of the major unknowns in terms of distribution, genetic structure, and ecological preferences of the *Leptidea* triplet in Europe. We present new molecular and morphological data from a detailed survey of the region. Our DNA analysis suggests that the Balkan populations belong only to *L. sinapis* and *L. juvernica*, and that they are not genetically differentiated from other mainland populations. The distribution data reveal that *L. sinapis* is a widespread habitat generalist, while *L. juvernica* exhibits a transition from a habitat generalist in the western Balkans to a localised habitat specialist, confined to humid mountain habitats, in the east. The morphometric analysis of male genitalia and a comparison to data from other parts of the species' ranges suggest an interesting mosaic of regional morphological variation that is likely linked to environmental and ecological factors. We also demonstrate the usefulness of the genitalia scaling relations for analysing the morphological variation and discuss the implications for species identification.

Keywords: cryptic species, genitalia morphology, variation, allometry, DNA

1. Introduction

The Wood White butterfly *Leptidea sinapis* (Linnaeus, 1758) and its two sibling species (*Leptidea reali* Reissinger, 1989 and *Leptidea juvernica* Williams, 1946) represent a fascinating example of the intricate relations developing within a cryptic species complex. Remarkably, in spite of more than two decades of extensive studies, the full taxonomic diversity of the complex was only recently brought into light. The existence of three well differentiated species was first revealed through molecular, karyological and morphological analyses (Dincă et al., 2011) and later reinforced by demonstrating reproductive isolation due to female mate choice (Dincă et al., 2013). The three species are virtually indistinguishable externally and two of them (*L. reali* and *L. juvernica*) also have extremely similar genitalia (Dincă et al., 2011). Because of its specific patterns of distribution and ecological interaction, the *Leptidea* triplet is an excellent model system for studying the origin and evolution of cryptic species, as well as their ecological interaction in conditions of sympatry (Dincă et al., 2013; Friberg et al., 2013). In addition, due to pronounced karyotype variability, this triplet (particularly *L. sinapis*) emerges as a promising model to study speciation linked to chromosomal rearrangements and sex determination systems (Dincă et al., 2011; Lukhtanov et al., 2011; Šíchová et al., 2015).

Of the three species, *L. sinapis* is the one with the widest distribution ranging from western Spain and Ireland to eastern Kazakhstan (Dincă et al., 2011, 2013) and further east to lake Baikal (Sinev, 2008). The other two species have allopatric distributions, but both are known to occur in sympatry with *L. sinapis* in parts of their respective ranges (Dincă et al., 2013). *Leptidea reali* occurs in Western Europe (Spain, southern France, and Italy) whereas *L. juvernica* extends from Ireland and France to eastern Kazakhstan (Dincă et al., 2011, 2013), Tian Shan in Kyrgyzstan, northwestern China (Bolshakov, 2006) and Republic of Tuva (Russia) (Sinev, 2008). Old (pre-2011) reports for *L. reali* from Central Europe eastwards therefore refer to *L. juvernica*, although the situation has not been fully clarified in the Balkans (see below).

Studies of the habitat preferences of *Leptidea* populations have revealed a variable niche specialization in different parts of the species' ranges. The same species can be a habitat generalist in some regions and a habitat specialist in others, where the sympatric sibling species may be a habitat generalist (Friberg et al., 2013). For example, in Sweden *L. juvernica* is a habitat specialist while *L. sinapis* is a generalist (Friberg et al., 2008); in Poland both species are widespread generalists (Sachanowicz et al., 2011); in the Czech Republic *L. sinapis* is the habitat specialist (Beneš et al., 2003); and in Slovenia both species are widespread with partial habitat segregation observed near the Adriatic Sea where *L. juvernica* is confined to humid habitats (Verovnik et al., 2012). Documenting this geographical variation can provide essential information for understanding the causes of niche specialization and their relation to the microevolutionary processes that eventually produce sibling species (Dincă et al., 2011; Friberg et al., 2013).

The Balkan Peninsula remains relatively poorly studied in terms of species composition, distribution, ecology, and genetic structure of the *Leptidea* triplet. Data from this region are scarce and sporadic. Some authors assumed that all unverified reports from the region belong to *L. sinapis* (Abadjiev, 2001; Pamperis, 2009). A series of other studies based on genitalia morphology or DNA data have been published, with data for Croatia (Lorkovic, 1993; Mazel and Leestmans, 1999; Dincă et al., 2011, 2013), Bosnia-Herzegovina (Lelo, 2007, 2010), Bulgaria (Kristal and Nässig, 1996; Mazel and Leestmans, 1999; Dincă et al., 2011, 2013), Greece (Kristal and Nässig, 1996; Mazel and Leestmans, 1999; Dincă et al., 2011, 2013; Coutsis, 2013), and Romania (Rákosy, 1996; Dincă et al., 2011, 2013). *Leptidea reali* sensu lato has also been reported from several localities in Serbia and Montenegro (Jakšić, 1999; Jakšić and Ristić, 1999); however, these records may represent erroneous identifications because the figured male genitalia (Jakšić and Ristić, 1999) do not correspond to the *L. reali-juvernica* type. All these works reported isolated findings that were not a result of a systematic survey. Moreover, since *L. reali* and *L. juvernica* appear to have overlapping genitalia morphology, even studies using this character have limited resolution (being able to separate only *L. sinapis*) in the absence of DNA data. Notably, with the exception of the northwestern limit (Slovenia), all available DNA-based data from the Balkans refer to *L. sinapis*. It is thus necessary to clarify whether previous records of *L. reali* from this area actually refer to *L. juvernica* or not. The known distribution of *L. reali* is largely restricted to two of the three main European peninsulas (Iberia and Italy), and the possibility that populations of this species also occur in the Balkans cannot be discarded without verification.

The degree of intraspecific (including regional) vs. interspecific morphological variation of the *Leptidea* triplet presents another largely overlooked problem. Dincă et al. (2011) have noted the almost complete overlap in genitalia morphology of *L. reali* and *L. juvernica*, and the overlap in the size of genitalia features between *L. juvernica* and *L. sinapis* has been highlighted in other studies (Freese and Fiedler, 2004; Verovnik and Glogovčan 2007; Coutsis 2013), but has not been quantitatively studied. In addition, an overview of published data by Sachanowicz (2013) has indicated that the ranges of variation of genitalia measurements may differ significantly between regions. The extent to which this variation can affect the accuracy of discrimination among the species has not been investigated in detail, in spite of its potentially profound implications on our ability to use genitalia morphology to reliably identify the sympatric species pairs and trace their regional distribution.

The objective of the present study is to partially fill the existing gap in our knowledge of the *Leptidea* triplet by establishing what species occur in the Balkan Peninsula and by carrying out a systematic study of their distribution, ecology, genetic structure, and morphological variation. A detailed morphometric analysis combined with molecular data is used for identification, and the regional morphological and genetic variation of the Balkan populations is placed in a broader context by a comparison with published data from other parts of the species' ranges. Habitat preferences and conservation aspects are discussed in the context of the known variable niche specialization of *Leptidea*.

2. Material and methods

2.1. DNA analyses

We sequenced 655-bp of the mitochondrial cytochrome *c* oxidase subunit I gene (COI) for 16 *Leptidea* specimens, 13 of which originated from the Balkans (Supplementary Table S1). Total genomic DNA was extracted using Chelex 100 resin, 100–200 mesh, sodium form (Biorad), under the following protocol: one leg was removed and introduced into 100 μL of Chelex 10% and 5 μL of Proteinase K (20 mg/mL) were added. The samples were incubated overnight at 55ºC and were subsequently incubated at 100ºC for 15 minutes. Samples were then centrifuged for 10 s at 3000 rpm. A 655-bp fragment at the 5´ end of COI was amplified by polymerase chain reaction using the primers LepFb (5´-ATTCAACCAATCATAAAGATATTGGAAC-3´) (a slightly modified version of LepF) and LepR (5´-TAAACTTCTGGATGTCCAAAAAATCA-3´) (Hajibabaei et al., 2006). Double-stranded DNA was amplified in 25-μL volume reactions containing: 14.4 μL autoclaved Milli-Q water, 5 μL 5x buffer, 2 μL 25 mM MgCl$_2$, 0.5 μL 10 mM dNTPs, 0.5 μL of each primer (10 μM), 0.1 μL Taq DNA Polymerase (Promega, 5U/ μL) and 2 μL of extracted DNA. The typical thermal cycling profile was: first denaturation at 92ºC for 60 s, followed by five cycles of 92ºC for 15 s, 48ºC for 45 s and 62ºC for 150 s, and then by 35 cycles of 92ºC for 15 s, 52ºC for 45 s and 62ºC for 150 s and a final

extension at 62ºC for 420 s. PCR products were purified and sequenced by Macrogen Inc. Sequences were edited and aligned using GENEIOUS PRO 4.7.5 created by Biomatters (http://www.geneious.com/).

To place the genetic structure of the Balkan populations in a broader context, we combined our data with the 410 COI sequences of *Leptidea sinapis, L. reali*, and *L. juvernica* from Dincă et al. (2013). This combined dataset included 55 *Leptidea* specimens from the Balkans (Supplementary Table S1). Eight sequences of *L. morsei*, *L. amurensis*, *L. lactea*, and *L. duponcheli* were used as outgroup. To assess the identification of the samples sequenced and the intraspecific genetic variability, MEGA6 (Tamura et al., 2013) was used to infer a neighbour-joining (NJ) tree based on *p*-distance and pairwise deletion, and node supports were assessed through 1000 bootstrap replicates. For practical purposes, the NJ tree was based on unique *L. sinapis, L. reali*, and *L. juvernica* haplotypes that were obtained using the program TCS 1.21 (Clement et al., 2000) (Supplementary Table S1). Thus, the final alignment included 46 COI haplotypes (Supplementary Table S1) and eight outgroup sequences and was 658 bp long.

2.2. Morphological analyses

A total of 249 *Leptidea* specimens (214 ♂♂, 35 ♀♀) collected in the Balkan Peninsula were studied, including 86 specimens obtained from the collection of the National Museum of Natural History (NMNH) in Sofia, Bulgaria. The specimen data are listed in Supplementary Table S2. Abdomens were macerated in 10% KOH for 15 min at 90ºC and dissected under a stereomicroscope. Genitalia were photographed in a drop of distilled water in natural position without compressing or flattening them. Special care was taken to arrange the measured linear structures parallel to the focal plane of the microscope in order to minimize the measurement error. A 2-megapixel USB camera was used with a 4x objective, ensuring a uniform resolution of 711 pixels/mm across the frame measured with a calibration slide (0.01 mm divisions). The mean accuracy of measurements of individual dimensions was determined to be 0.015 mm, associated mainly with the uncertainty in determining the limits of individual features. The following dimensions were measured from the digital photographs: aedeagus length (AL), vinculum/valva width (VW), saccus length (SL), and uncus length (UL) for males and ductus bursae length (DBL) for females. Typical male and female genitalia are shown in Fig. 1 together with the relevant dimensions. The measured variables are the most commonly used in the literature and therefore have the advantage of being directly comparable between different data sets (see e.g. Fumi, 2008; Dincă et al., 2011). However such comparisons should be done carefully as some authors have used slightly different variables, e.g. the ventral length of the genital capsule instead of VW (Sachanowicz, 2013).

Statistical analyses of the morphometric data set including Shapiro-Wilk's normality tests, principal component analysis, and linear discriminant analysis with leave-one-out cross-validation were carried out using the statistical software R (R Core Team, 2013). Male specimens were identified by model-based clustering analysis with finite Gaussian mixture models as implemented in the R package mclust (Fraley and Raftery, 2002). The best model was selected based on the Bayesian Information Criterion (BIC). The extent to which the Balkan populations, as a whole, are morphologically distinct from other populations was estimated for each measured variable using a two-tailed t-test with unequal variances. The data set of Dincă et al. (2011) was used for comparison as it samples a significant part of the species ranges and hence is assumed to represent the geographically averaged morphotypes for each species. From here on this data set is referred to as DLTV, abbreviated from the names of its authors.

Since the detailed study of the DLTV dataset and the performed t-tests revealed significant regional differentiation in the size of the genitalia, we carried out a second, more detailed analysis on a combined dataset including our eastern Balkan sample and the DLTV data (including six specimens collected in the Balkans). The regional variation in shape and size of male genitalia was analysed by studying the scaling relations of genitalia dimensions. Published averages for regional samples were included for comparison, when available in a compatible form. The results of such analyses depend significantly on the choice of a "size" variable or combination of variables (e.g. Mosimann, 1970; Jungers et al., 1995). Previously, a geometric mean of all measurements has been used as a size variable in *Leptidea* (Fumi, 2008), but the use of VW has also been suggested (Vila et al., 2003; Dincă et al., 2011). We used VW as a size variable since it represents better the overall size of the genital capsule, while the variation in AL and SL (at constant VW) is more intuitively interpreted as a variation in shape. In a comparative context, this choice also has the advantage that the interspecific differences (expressed mostly in AL and SL) can be separated from the intraspecific size variation. As all variables were normally distributed, using a logarithmic transformation was not necessary (Jungers et al., 1995). A test for possible correlations between the size variation and the genetic differentiation was carried out using Blomberg et al.'s (2003) K-statistic on the averaged values of the size variable (VW) for each haplotype and the phylogenetic tree obtained from the DNA analysis.

To investigate the effect of the observed variation on the reliability of discrimination between species, linear discriminant functions (DFs) were obtained for different sets of variables in the combined dataset using linear discriminant analysis. The performance of each DF was estimated from the probability for misclassification assuming normally distributed discriminant scores, estimated means and variances of each distribution from the combined dataset, and equal priors. For simplicity, all coefficients were scaled to obtain a coefficient of one for the first variable and the DF was cast in a form that gives a threshold of zero: negative values of the DF correspond to *L. sinapis* and positive

values to *L. juvernica*.

3. Results

3.1. DNA data

The DNA analyses (Fig. 2, Supplementary Table S1) indicated that 41 of the sequenced specimens from the Balkans belong to *L. sinapis* (Bosnia and Herzegovina, Bulgaria, Croatia, Greece, Macedonia, Romania, Serbia, and Slovenia), while 14 belong to *L. juvernica* (Slovenia, southwestern Bulgaria, and northern Greece). No *L. reali* were identified among the samples analysed. DNA-based identifications were always congruent with the morphology of the genitalia for the analysed specimens (Supplementary Table S2).

The Balkan *L. sinapis* were represented by seven different haplotypes out of 25 in the entire dataset. Only two haplotypes of *L. sinapis* were detected exclusively in the Balkans (hs20 and hs22), both differentiated by only one substitution. The Balkan *L. juvernica* comprise five different haplotypes out of 14 in the dataset. Nine of the 14 sequenced *L. juvernica* belong to the most widespread hj1 haplotype, which occurs throughout the species' range from Germany and the Czech Republic to Russia and Kazakhstan. All five Balkan haplotypes were detected in Slovenia and two haplotypes (hj13 and hj14) were detected only in the Balkans, but, like in *L. sinapis*, they are both differentiated by only one substitution. No clear geographic pattern of intraspecific genetic differentiation was detected based on the analysed Balkan samples.

3.2. Identification

Females. Previous morphometric studies of large samples have shown that the distribution of DBL in females is clearly bimodal without any overlap between the two species (Freese and Fiedler, 2004; Cupedo and Hoen, 2006; Tsvetkov, 2007; Sachanowicz, 2013). Our sample agrees with this observation, thus allowing the unambiguous identification of females (Table 1).

Males. Males were identified by model-based clustering analysis of the dataset consisting of the four genitalia variables. The model with best statistical support was VEV (ellipsoidal distribution with variable volume, equal shape and variable orientation) with two groups. The four models with ellipsoidal distribution and two groups were overwhelmingly supported by the data (BIC model weights 0.74, 0.14, 0.12, and 0.0004) and resulted in identical classifications. To examine the robustness of the obtained classification, linear discriminant analysis was carried out for all male specimens and posterior probabilities were calculated using leave-one-out cross-validation with equal priors (Kuhn and Johnson, 2013: 70). The results confirmed the initial identification obtained from the cluster analysis, with all posterior probabilities larger than 0.9995. The results from principal components (PC) analysis shown in Fig. 3 also demonstrated that all specimens are readily and unambiguously separated into two groups based on PC1 and PC2 (accounting for 58% and 30% of the variation), in complete agreement with the identification obtained from the clustering model.

3.3. Variation

Summary statistics of all measured genitalia variables in male and female specimens are shown in Table 1. The Shapiro-Wilk test confirmed that all variables have a normal distribution ($p > 0.05$). The variation ranges for AL and SL are greater in *L. juvernica* than in *L. sinapis*, in agreement with the results of previous studies (Sachanowicz, 2013). There is no overlap in the values of those variables between two species; the same is true for the DLTV dataset.

The t-test comparisons of *L. sinapis* to the DLTV dataset showed that the distributions of all variables were virtually identical to the DLTV data (p-values listed in Table 1). The p-value for UL ($p = 0.041$) was only marginally lower than the nominally used threshold of 0.05; therefore we conclude that Balkan *L. sinapis*, as a whole, are remarkably similar to the "range average" genital morphology. The situation was completely different for *L. juvernica*, where the average values of AL, SL, and VW for the Balkan sample were all significantly larger ($p \leq 0.001$) than the corresponding values in the DLTV data set.

3.4. Morphological variation and identification in a broad geographical context

The scaling of AL, SL, and UL with the overall size of genitalia (represented by VW) in each species was fitted using linear regression models on the combined dataset (present sample and DLTV). The resulting regression coefficients and the associated standard errors, p-values, and R^2 are shown in Table 2. All scaling relations are allometric, as indicated by the non-zero intercepts; thus, the shape of the genitalia changes simultaneously with their size (contrary to the isometric case, in which all dimensions are proportional to size and hence have zero intercepts). T-tests on the

regression residuals confirmed that, after removing the size dependence, no significant difference remained between the Balkan sample and the DLTV dataset ($p > 0.05$ for all three variables).

Further details on the regional variation of genitalia were obtained by analysing the average dimensions by country from the DLTV dataset and other available sources (Fumi, 2008; Sachanowicz, 2013). For the comparison, the Balkan specimens from the combined dataset were split into two regions: eastern Balkans (Bulgaria, Turkey, northeastern Greece, southeastern Romania), and western Balkans (all remaining territory, including Slovenia). The results are shown in Fig. 4 and demonstrate that the country/regional averages (including those from other sources) conform very closely to the obtained allometric scaling relations for both *L. sinapis* and *L. juvernica*. We note that not all data points are truly representative for the regional averages as the number of specimens per country varies (for the DLTV dataset $5 \leq N \leq 20$, except for Irish *L. juvernica*: $N = 2$, but see below). In addition, even large regional samples may cover only a small subset of the environmental and ecological gradients present in a region, possibly introducing sample bias. However, it is remarkable that all regional averages agree closely with our linear regression fit. From the limited available data for *L. reali* (not shown in the figure), it appears that the scaling relations for this species are very similar to the ones for *L. juvernica*.

Eastern-Balkan and Irish *L. juvernica*, as well as Italian and Irish *L. sinapis*, emerged as the most significantly differentiated regional samples (all of these being larger than the average for the respective species). The larger average size of Irish populations of both species is corroborated by published data (Nelson et al., 2011). Remarkably, the Italian sample of *L. sinapis* (Fumi, 2008) showed significantly larger dimensions than the DLTV subset of 13 specimens from Italy. This is possibly due to sample bias (see above), although different calibration and measurement procedures may also have played a role. Western-Balkan *L. sinapis* were clearly differentiated from eastern populations and appeared to have the smallest genitalia of all regional samples, in agreement with published data from Croatia (Lorković, 1993). No significant phylogenetic signal was detected in the size (VW) variation of male genitalia ($K = 0.174$, $p = 0.195$, 1000 randomization rounds).

The DF between *L. juvernica* and *L. sinapis*, obtained for the best performing combinations of variables, are shown in Table 3. The full DF containing all four genitalia variables had an error rate of about one in 2000 specimens, thus demonstrating that the identification based on the full shape and size of the genitalia is very reliable. From all two-variable DFs, the one based on AL and VW had the best performance (1/1250). The DF obtained using an average between the two AL scaling relations (Fig. 4a) was essentially identical, which is to be expected given that both functions are linear and are based on the same variables. All other combinations of variables had inferior performance.

3.5. Distribution and phenology

A map of the distribution of *L. sinapis* and *L. juvernica* in the Balkan Peninsula, based on the studied specimens (Supplementary Tables 1 and 2) is shown in Fig. 5. Our findings indicate that *L. sinapis* is widely distributed in the region, found at 134 distinct localities at altitudes ranging from sea level to 1900 m and spanning a variety of habitats from xerothermophilous submediterranean scrub to mesophilous forest clearings and mountain pastures near the tree line. In contrast, *L. juvernica* was found at only 16 localities. While at the western extreme of the region (Slovenia) the species was found at low to intermediate altitudes (Verovnik and Glogovčan, 2007; Verovnik et al., 2012), all our records from the thoroughly sampled eastern part of the Peninsula came from high mountains (Vitosha, Rila, Pirin, and Rodopi Mts., all with highest peaks exceeding 2100 m). Here, the species was observed at altitudes ranging from 700 m (Vitosha: Knyazhevo) to 1900 m (Vitosha: Goli Vrah; Pirin: "Kameniti dupki") in coniferous and mixed woodland, forest meadows and clearings, and high-altitude meadows adjacent to the tree line. The two species were found to co-occur within the range of *L. juvernica*, with syntopy observed at 11 localities (69%). This fraction is likely underestimated as not all localities have been thoroughly sampled.

An offset in phenology of up to two weeks has been observed between allotopic *L. sinapis* and *L. juvernica* populations in Sweden (Friberg et al., 2008). This has been attributed to differences in developmental conditions between the meadow and forest habitats occupied by the two species, but has also been linked to their different responses to larval diet and temperature during development (Friberg and Wiklund, 2010). In addition, allochrony could be developed in sympatric populations as a means for offsetting some of the survival cost associated with heterospecific courtships (Friberg et al., 2013). To estimate the level of synchronicity in the two species, we analysed the phenology of the eastern-Balkan subset of our sample, which represents a thoroughly sampled and relatively localised area. A breakdown by date and altitudinal belt is shown in Fig. 6. Both species appear to have similar phenology; the adults fly in two slightly overlapping generations, from April–May and June–July to May–June and July–August depending on altitude. There is also a possible third generation in August–September in warm climates (see also Pamperis, 2009). Our data are not sufficiently detailed to reveal small offsets in population peaks; however, within the available resolution, both our field observations and historical collection data indicate that the two species fly synchronously where they co-occur.

4. Discussion

4.1. DNA

Previous studies have suggested that *L. juvernica* and *L. reali* are synmorphic species that can be differentiated from *L. sinapis* through male or female genitalia examination, but can be separated from each other only by DNA data (Dincă et al., 2011). The available molecular data have shown that *L. reali* is confined to the northern half of Iberia, parts of southern France and at least central Italy, while *L. juvernica* replaces it in the rest of the Eurasian range (Dincă et al., 2011, 2013). However, the situation in the Balkans has, until now, remained unclear: although reports of specimens with genitalia of *L. reali* type (as published before 2011) do exist (e.g. Lorkovic, 1993; Kristal and Nässig, 1996; Mazel and Leestmans, 1999), records based on molecular data are available only for the western extreme of the region (Slovenia; see Verovnik and Glogovčan, 2007; Dincă et al., 2011, 2013). Given the known distribution of *L. reali*, the hypothesis that this species has a distribution restricted to the three main Southern-European peninsulas (and glacial refugia) would appear plausible. However, our DNA analyses indicate that this is unlikely, since all 55 analysed specimens from the Balkans (including 13 produced by this study) belong to *L. sinapis* and *L. juvernica* (Fig. 2, Supplementary Table S1). Moreover, we have shown that the Balkan populations of *L. sinapis* and *L. juvernica* do not represent separately evolving units and that, among the studied populations, only the Irish *L. juvernica* show a geographically-correlated genetic differentiation (Dincă et al., 2013) (Fig. 2).

4.2. Morphological variation

The morphological variation of *L. sinapis* and *L. juvernica* is difficult to assess rigorously because of the uneven sample sizes and different calibration and measurement procedures of published data. The use of the DLTV dataset (Dincă et al., 2011) as a reference to which to compare regional samples has allowed us to carry out this comparison and provides the advantage of a broad geographical coverage for each species. For *L. sinapis*, a comparison of the DLTV averages to published data from Bosnia-Herzegovina (Lelo, 2002), Germany (Freese and Fiedler, 2004), and Poland (Sachanowicz, 2013) shows a general agreement in the range of variation of different genitalia dimensions. This is corroborated by a more rigorous comparison to our Balkan dataset (Table 2) and supports the hypothesis that the genitalia morphology of *L. sinapis* is relatively uniform across its range (Lukhtanov et al., 2011). However, published data from Italy (Fumi, 2008) and Ireland (Nelson et al., 2011) suggest that these regional populations have significantly larger genitalia (by 5–8%) compared to the DLTV averages. For *L. juvernica*, our morphometric analysis shows statistically significant differences in the size of the male genitalia in the Balkan sample compared to the DLTV average (Table 2). Larger than average male genitalia in this species have been observed also in Ireland (Nelson et al., 2011), whereas in Poland (Sachanowicz, 2013) the reported dimensions are similar to the DLTV averages. This suggests an interesting mosaic of regional morphological variation and highlights the need for a more detailed analysis of the variability of both species.

A general framework for such analysis is provided by the scaling relations presented in Table 2 and Fig. 4. It is clear that, although there are significant differences in the average genitalia dimensions between different populations, all available regional samples follow closely the same scaling relations. Sample bias and/or different calibration and measurement procedures employed in different studies could influence the results to some extent. Still, the observed regional differentiation in the DLTV dataset, the results of the present study, and the close agreement of published data with the obtained scaling relations are a strong indication that the detected morphological variation of each species across their respective ranges is real. The causes of the observed regional differences are not clear; however, we detected no correlation between genetic differentiation and size variation. Thus, environmental and ecological factors such as climatic differences, habitat and host-plant specialisation are likely to play a leading role, as suggested also by Tzvetkov (2007). One documented effect related to the scaling of genitalia is the size difference between different generations that has been observed in both species (Schmitz, 2007; Sachanowicz, 2011). It appears, however, that the size variation between generations is much smaller than the observed regional variation: e.g., the VW differences reported by Sachanowicz (2011) were 0.03 mm in *L. sinapis* and 0.01 mm in *L. juvernica*, whereas the maximum regional differences in Fig. 4 are 0.17 mm and 0.05 mm, respectively. A preliminary analysis of both published data and our Balkan dataset shows that the variation of genitalia size between generations is well described by the scaling relations in Table 2 and Fig. 4. This indicates that generation is an additional factor that is responsible in part for the size variation, but it cannot explain it entirely. The interaction between sibling species is another factor that could possibly increase the size differentiation between sympatric populations, for instance as observed in giant rhinoceros beetles *Chalcosoma* sp. (Kawano, 2002). Further studies correlating the observed size differences with climatic variables, generation, habitat and host-plant specialization, and overall species distribution of the *Leptidea* triplet are needed to elucidate this intriguing pattern.

4.3. Identification

The scaling relations in Fig. 4 have direct implications for species identification. As evident from the PCA results (Fig. 3), AL and SL give the best discrimination between *L. sinapis* and *L. juvernica*. This was recognized in early studies;

however, the ranges of individual variables can overlap between the two species (e.g. Freese and Fiedler, 2004; Verovnik and Glogovčan, 2007; Coutsis, 2013). This circumstance renders the identification based only on those two variables less reliable (Fumi, 2008), a fact also indicated by the large error rate of the corresponding DF in Table 3. Including VW can greatly improve the identification success rate by effectively eliminating the part of the variation that is due to overall size of the genitalia. Due to the allometric scaling, however, size accounts for a significant part of the intraspecific as well as interspecific variation. Hence, using ratios such as AL/VW and SL/VW, as previously suggested (Vila et al., 2003; Dincă et al., 2011), eliminates part of the shape information and thus does not improve the discrimination accuracy compared to using only AL and SL (Table 3). On the other hand, full discriminant functions such as the one given by Fumi (2008) have theoretically the best performance. However, they are usually fitted to regional datasets with limited geographical coverage and are mathematically complex, which obscures their interpretation and limits their usefulness.

As a compromise between simplicity and performance, we recommend the use of the DF based on the scaling relation of AL (shown in bold in Table 3) for identification. It uses only AL and VW measurements, has the second-best error rate of less than 1/1250, and is based on the observed scaling of genitalia over a wide geographic region, thus largely eliminating the effect of the size variation. This makes it ideal for identifying single specimens without the need of a large comparison sample. In addition, this DF has a simple geometric interpretation: its sign corresponds to specimens below (negative sign) or above (positive sign) the dashed line shown in Fig. 4a. Our numerical estimates are based on discrimination between *L. sinapis* and *L. juvernica*; however, given the similarity of the latter species' genitalia morphology with *L. reali*, it appears the DF should perform similarly well for identifying the *L. reali/L. sinapis* species pair in Western Europe.

4.4. Distribution and ecology

Recent research has revealed the importance of habitat preferences and generalist/specialist relations between the two competing species in different parts of their range. Our results indicate that in the Balkan Peninsula *L. sinapis* is a widespread habitat generalist ranging from the driest and warmest parts of the region to cool mountain habitats near the tree line. In contrast, *L. juvernica* shows a geographically differentiated pattern of ecological specialization. While at the westernmost limit of the studied region it is a habitat generalist (Verovnik et al., 2012), in the eastern Balkans this species is a local habitat specialist with a strong preference for mesophilous mountain habitats, and we have not recorded it at altitudes below 700 m. It is worth noting that no genetic divergence was detected between the eastern and western *L. juvernica* populations, which share haplotype hj1. Our observations confirm that the two species coexist where their ranges overlap; in addition *L. juvernica* has been observed to be locally abundant in its preferred ecological niche, which may be essential for its survival in competition with *L. sinapis* (Friberg et al., 2013).

The outlined distribution and habitat preferences are corroborated by other published records based on genitalia or DNA identification (Mazel and Leestmans, 1999; Dincă et al., 2011, 2013; Coutsis, 2013). In particular, the mountain distribution of *L. juvernica* and its syntopy with *L. sinapis* are corroborated by published records of both species from Rila Mts.: "Bameken" [sic, recte Belmeken] (Mazel and Leestmans, 1999) and from the Greek part of Rodopi Mts. (Pamperis, 2009; Coutsis, 2013; J. Coutsis, pers. comm.). The predominantly mountain distribution in the southern limit of the species range suggests that *L. juvernica* may have a lower tolerance to warm and dry climates than *L. sinapis*. A similar distribution pattern has been observed for *L. reali/juvernica* in Southwestern Europe (Vila et al., 2003; Amiet, 2004).

The data from the central-western parts of the Balkan Peninsula are insufficient to determine whether the observed transition in the ecological preferences of *L. juvernica* from a habitat generalist in the west to a habitat specialist in the east happens gradually or abruptly. Several published records from Bosnia and Herzegovina (Lelo, 2007) suggest that *L. juvernica* occurs there at low to intermediate altitudes (200–800 m), much like in Slovenia; however it appears to be much more localised than the widespread *L. sinapis* (Lelo, 2010). These short reports, however, are in need of confirmation, since they give no details on the identification method and no precise specimen and locality data. Further studies with thorough sampling in this part of the Balkans are needed to trace the transition and help elucidate the correlation of the observed habitat specialization to climatic factors.

Two previous *L. juvernica* records from the southeastern part of the region deserve further comment: "Bulgaria, Jasnà Poljana" [Yasna Polyana village in southeastern Bulgaria near Burgas] and "Greece, Alexandrupolis" (Kristal and Nässig, 1996). Both localities are characterized by a warm and relatively dry climate and seem to represent a different climatic niche than what is suggested by our data. Unfortunately the exact localities are difficult to determine as the original report is based on loan specimens from different collections (Kristal and Nässig, 1995; W. Nässig, pers. comm.). In an effort to corroborate the first locality, we examined 11 specimens from southeastern Bulgaria from the NMNH collection, all of which turned out to be *L. sinapis*. In fact, no *L. juvernica* was detected among the 131 specimens studied that originated from regions in the eastern Balkans outside the Vitosha, Rila, Pirin, and Rodopi Mts. While the two records in question could come from isolated local populations that have adapted to a secondary ecological niche, we recommend that these reports are treated with caution until they are confirmed by further studies.

4.5. Conservation

It is clear from the current distribution and past and recent observations that *L. sinapis* is a common and widespread species that adapts well to both natural and anthropogenically created and maintained habitats, and is therefore of no conservation concern in the region. Unlike the former, *L. juvernica* is more local in the eastern Balkans (Bulgaria and Greece) and therefore more vulnerable to habitat loss and local extinction, but there is no evidence that its populations have declined. The specimens examined in the course of this study were collected in a time span of 109 years (1906–2014) and our data confirm the continuing presence of *L. juvernica* near some of the localities where it has been found in the beginning of the 20th century in the Vitosha and Rodopi Mts. Its habitats in the eastern Balkans are not threatened and many of the reported localities are included within protected areas (Vitosha Nature Park, Rila and Pirin National Parks in Bulgaria; Natura 2000 sites in Bulgaria and Greece). Thus, the evaluation by IUCN criteria A (population decline) and B (range decline) (IUCN Standards and Petitions Subcommittee, 2013) based on the available data suggests that both *L. sinapis* and *L. juvernica* are not threatened (IUCN category: Least Concern).

5. Conclusion

Until now, the Balkan Peninsula remained one of the major unknowns in the distribution, genetic structure, and ecological preferences of the *Leptidea* triplet of sibling species in Europe. Our molecular data indicate that the *Leptidea* triplet is most likely represented in the Balkans only by *L. sinapis* and *L. juvernica*, while *L. reali* is confined to Iberia, southern France and Italy. We also show that the Balkan populations of *L. sinapis* and *L. juvernica* are not genetically differentiated from other mainland populations.

Our systematic survey of the territory of the Balkan Peninsula sheds light on the distribution and habitat preferences of *L. sinapis* and *L. juvernica* in the region. The results reveal that in the Balkans, *L. sinapis* is a widespread habitat generalist occurring at altitudes from sea level to ca. 2000 m. In contrast, *L. juvernica* exhibits a transition from a habitat generalist in the west to a localised habitat specialist in the east, where it is confined to more humid mountain habitats (forests, meadows, and forest clearings) at 700–2000 m a.s.l. It is suggested that the observed niche specialisation is at least partially due to the different climatic preferences of the two species. Further studies are needed to trace the transition in detail and elucidate the correlation between habitat specialisation and environmental (including climatic) factors.

A morphometric analysis of the male genitalia shows a mosaic of regional morphological variation. As a whole, the Balkan populations of *L. sinapis* are morphologically similar to the range averages, while *L. juvernica* has significantly larger genitalia than populations from other regions. We show that the scaling relations of genitalia dimensions with respect to the vinculum/valva width (VW) provide a useful framework for describing the entire range of morphological variation observed in both species, and use them to derive a simple and efficient discriminant function suitable for identifying single specimens. No correlation is observed between the DNA data and the morphological variation in either species, suggesting that the observed differences are likely a result of environmental and ecological factors and highlighting the need for further studies in order to understand this intriguing result.

Acknowledgements

References

Abadjiev, S.P., 2001. An Atlas of the Distribution of the Butterflies in Bulgaria (Lepidoptera: Hesperioidea & Papilionoidea). Pensoft Publishers, Sofia–Moscow, 335 pp.

Amiet, J.L., 2004. Séparation des niches écologiques chez deux espèces jumelles sympatriques de *Leptidea* (Lepidoptera, Pieridae). Revue d'écologie (la Terre et la Vie) 59, 433–452.

Beneš, J., Konvička, M., Vrabec, V., Zámečník, J., 2003. Do the sibling species of small whites, *Leptidea sinapis* and *L. reali* (Lepidoptera, Pieridae) differ in habitat preferences? Biologia 58, 943–951.

Blomberg, S.P., Garland Jr., T., Ives, A.R., 2003. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. Evolution 57, 717–745.

Bolshakov, L.V., 2006. New subspecies of *Leptidea reali* Reissinger, 1989 (Lepidoptera: Pieridae) from the mountain regions of Middle Asia. Eversmannia 5, 6–10.

Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9,** 1657–1660.

Coutsis, J.G., 2013. *Leptidea sinapis* and *Leptidea reali* (Lepidoptera: Pieridae): at what point does the first one of the two end, and the other one begin? Phegea 41, 19–20.

Cupedo, F., Hoen, F.W., 2002. *Leptidea sinapis* and *Leptidea reali* (Lepidoptera: Pieridae) in The Netherlands. Entomologische Berichten 66, 118–123.

Dincă, V., Lukhtanov, V.A., Talavera, G., Vila, R., 2011. Unexpected layers of cryptic diversity in wood white *Leptidea* butterflies. Nature Communications 2, 324.

Dincă, V., Wiklund, C., Lukhtanov, V.A., Kodandaramaiah, U., Norén, K., Dapporto, L., Wahlberg, N., Vila, R., Friberg, M., 2013. Reproductive isolation and patterns of genetic differentiation in a cryptic butterfly species complex. Journal of Evolutionary Biology 26, 2095–2106.

Fraley, C., Raftery, A.E., 2002. Model-based clustering, discriminant analysis and density estimation. Journal of the American Statistical Association 97, 611–631.

Freese, A., Fiedler, K., 2004. Unterscheidungsmerkmale von *Leptidea sinapis* (Linnaeus, 1758) und *Leptidea reali* Reissinger, 1989 (Lepidoptera, Pieridae) und ihre Eignung zur Artbestimmung. Nachrichten des entomologischen Vereins Apollo, N. F. 25, 65–77.

Friberg, M., Bergman, M., Kullberg, J., Wahlberg, N., Wiklund, C., 2008. Niche separation in space and time between two sympatric sister species—a case of ecological pleiotropy. Evolutionary Ecology 22, 1–18.

Friberg, M., Leimar, O., Wiklund, C., 2013. Heterospecific courtship, minority effects and niche separation between cryptic butterfly species. Journal of Evolutionary Biology 26, 971–979.

Friberg, M., Wiklund, C., 2010. Host-plant-induced larval decision-making in a habitat/host-plant generalist butterfly. Ecology 91, 15–21.

Fumi, M., 2008. Distinguishing between *Leptidea sinapis* and *L. reali* (Lepidoptera: Pieridae) using a morphometric approach: impact of measurement error on the discriminative characters. Zootaxa 1819, 40–54.

Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W., Hebert P.D.N., 2006. DNA barcodes distinguish species of tropical Lepidoptera. Proceedings of the National Academy of Sciences, USA 103, 968–971.

IUCN Standards and Petitions Subcommittee, 2013. Guidelines for Using the IUCN Red List Categories and Criteria. Version 10. Prepared by the Standards and Petitions Subcommittee. http://www.iucnredlist.org/documents/RedListGuidelines.pdf, downloaded on 17.06.2013.

Jakšić, P. 1999. The butterflies of Serbian part of Stara Planina Mt. (Lepidoptera: Hesperioidea and Papilionoidea). Nature protection 51, 69–84.

Jakšić, P., Ristić, G., 1999. New and rare species of Lepidoptera in Yugoslavia. Acta entomologica serbica 4, 63–74.

Jungers, W.L., Falsetti, A.B., Wall, C.E., 1995. Shape, relative size, and size-adjustments in morphometrics. Yearbook of Physical Anthropology 38: 137–161.

Kawano, K., 2002. Character displacement in giant rhinoceros beetles. The American Naturalist 159, 255–271.

Kristal, P.M., Nässig, W.A., 1996. *Leptidea reali* Reissinger 1989 auch in Deutchland und einigen anderen europäischen Ländern (Lepidoptera: Pieridae). Nachrichten des entomologischen Vereins Apollo, N. F. 16, 345–361.

Kuhn, M., Johnson, K., 2013. Applied predictive modeling. Springer, New York, xiii + 600 pp.

Lelo, S., 2002. Variation in exogenous and endogenous (genitalia) characteristics of butterflies of the species *Leptidea sinapis* Linnaeus, 1758 (Pieridae, Dismorphiinae) within populations from the area around Sarajevo. Natura Croatica 11, 293–319.

Lelo, S., 2007. Contribution to knowledge of the fauna of butterflies in Bosnia and Herzegovina. Acta entomologica serbica 12, 73–92.

Lelo, S., 2010. Rasprostranjenost gorušičinog bijelca, *Leptidea sinapis* (Linnaeus, 1758) (Insecta: Lepidoptera, Pieridae), u Bosni i Hercegovini. Prilozi fauni Bosne i Hercegovine, Sarajevo, 6, 1–9.

Lorkovic, Z., 1993. *Leptidea reali* Reissinger, 1989 (=lorkovicii Real, 1988), a new European species (Lepid., Pieridae). Natura Croatica 2, 1–26.

Lukhtanov, V.A., Dincă, V., Talavera, G., Vila, R., 2011. Unprecedented within-species chromosome number cline in the Wood White butterfly *Leptidea sinapis* and its significance for karyotype evolution and speciation. BMC Evolutionary Biology 11, 109.

Martin, J., Gilles, A., Descimon, H., 2003. Species concepts and sibling species: The case of *Leptidea sinapis* and *Leptidea reali*. In: Boggs, C. L., Watt, W. B., Ehrlich, P. R. (Eds.), *Butterflies: Ecology and Evolution Taking Flight*. Chicago University Press, pp. 459–476.

Mazel, R., Leestmans, R., 1999. Seconde contribution à l'étude des relations entre *Leptidea sinapis* Linné et *L. reali* Reissinger en France, Belgique et quelques autres contrées européennes (Lepidoptera: Pieridae). Linneana Belgica 17, 155–168.

Mosimann, J.E., 1970. Size allometry: Size and shape variables with characterizations of the lognormal and generalized Gamma distributions. Journal of the American Statistical Association 65, 930–945.

Nelson, B., Hughes, M., Bond, K., 2011. The distribution of *Leptidea sinapis* (Linnaeus, 1758) and *L. reali* Reissinger, 1989 (Lepidoptera: Pieridae) in Ireland. Entomologist's Gazette 62, 213–233.

Pamperis, L.N., 2009. The Butterflies of Greece, 2nd Ed. Athens: Editions Pamperis.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, http://www.R-project.org (accessed 31.12.2013).

Réal, P., 1988. Lépidoptères nouveaux principalement jurassiens. *Mém. Comité de Liaison Rech. Ecofaunist. Jura* **4,** 1–28.

Rákosy, L., 1996. *Leptidea reali* Reissinger, 1989 (Lepidoptera: Pieridae) specie nouă pentru fauna României. Bul. inf. Soc. lepid. rom. **7**, 171–177.

Sachanowicz, K., 2013. Separation possibilities and genital measurement variations in two cryptic species of European pierid butterflies, *Leptidea juvernica* Williams, 1946 and *L. sinapis* (Linnaeus, 1758). Zoology 116, 215–223.

Sachanowicz, K., Wower, A., Buszko, J., 2011. Past and present distribution of the cryptic species *Leptidea sinapis* and *L. reali* (Lepidoptera: Pieridae) in Poland and its implications for the conservation of these butterflies. European Journal of Entomology 108, 235–242.

Schmitz, O., 2007. Neueste Kenntnisse zur historischen und aktuellen Verbreitung von Leptidea sinapis (Linnaeus, 1758) und Leptidea reali Reissinger, 1989 (Lepidoptera, Pieridae) im Arbeitsgebiet der AG rheinisch-westfälischer Lepidopterologen. Entomol. Heute 19, 181–195.

Šíchová, J., Voleníková, A., Dincă, V., Nguyen, P., Vila, R., Sahara, K., Marec F., 2015. Dynamic karyotype evolution and unique sex determination systems in *Leptidea* wood white butterflies. BMC Evolutionary Biology 15, 89.

Sinev, S.Yu., 2008. Catalogue of the Lepidoptera of Russia. KMK Scientific Press, St. Petersburg–Moscow, 424 pp.

Székely, L., 2008. The Butterflies of Romania. C2 Design, Braşov.

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

Tsvetkov, E.V., 2007. On *Leptidea reali* Reissinger, 1989 (Lepidoptera: Pieridae) genitalia structure variability in Leningrad and Voronezh Areas. Eversmannia 11–12, 19–23.

Verovnik, R., Glogovčan, P., 2007. Morphological and molecular evidence of a possible hybrid zone of *Leptidea sinapis* and *L. reali* (Lepidoptera: Pieridae). European Journal of Entomology 104, 667–674.

Verovnik, R., Rebeusek, F., Jež, M., 2012. Atlas of butterflies of Slovenia. Centre for Cartography of Fauna and Flora, Miklavz na Dravskem Polju, 456 pp.

Vila, R., Viader, S., Jubany, J., 2003. *Leptidea sinapis* (Linnaeus, 1758) i *L. reali* (Reissinger, 1989): dues espècies "bessones" a Catalunya i Andorra (Lepidoptera: Pieridae). Butlletí de la Societat Catalana de Lepidopterología 90, 25–47.

Tables

Table 1. Summary statistics of genitalia dimensions for the Balkan sample. The variables are explained in Fig. 1. P-values from two-tailed t-tests are calculated with respect to the DLTV data set of Dincă et al. (2011); values showing significant differences ($p < 0.05$) are shown in bold.

Species		*L. sinapis*			*L. juvernica*		
sex	Var.	mean ± stdev (mm)	range (mm)	t-test DLTV	mean ± stdev (mm)	range (mm)	t-test DLTV
male	*n*	182		48	32		48
	AL	1.59 ± 0.07	1.39–1.76	0.87	2.06 ± 0.11	1.80–2.34	**< 10⁻³**
	SL	0.63 ± 0.05	0.49–0.76	0.70	0.91 ± 0.07	0.81–1.08	**0.001**
	VW	0.79 ± 0.05	0.63–0.93	0.93	0.82 ± 0.04	0.76–0.93	**< 10⁻⁴**
	UL	0.64 ± 0.04	0.50–0.79	**0.041**	0.64 ± 0.04	0.53–0.74	0.25
female	*n*	32			3		
	DBL	0.61 ± 0.04	0.51–0.67	—	1.12 ± 0.09	1.03–1.19	—

Table 2. Scaling relations of genitalia dimensions (AL, SL, and UL) with overall size (VW) for the combined Balkan and DLTV datasets. The calculated regression coefficients, standard errors, p-values, and coefficient of determination are listed for each regression fit.

	L. sinapis					*L. juvernica*				
Var.	Slope	P	Intercept (mm)	P	R^2	Slope	P	Intercept (mm)	P	R^2
AL	0.97±0.08	$< 10^{-4}$	0.83±0.06	$< 10^{-4}$	0.41	1.31±0.21	$< 10^{-4}$	0.98±0.17	$< 10^{-4}$	0.33
SL	0.35±0.07	$< 10^{-4}$	0.36±0.05	$< 10^{-4}$	0.10	0.55±0.16	10^{-3}	0.45±0.12	$< 10^{-3}$	0.14
UL	0.55±0.04	$< 10^{-4}$	0.21±0.03	$< 10^{-4}$	0.40	0.36±0.09	$< 10^{-4}$	0.35±0.07	$< 10^{-4}$	0.18

Table 3. Comparison of linear discriminant functions for the identification of *L. sinapis* and *L. juvernica*. The coefficients for all functions are calculated using linear discriminant analysis (LDA) except for one, which is derived from the scaling relations of AL (Table 3). Positive values of the discriminant functions correspond to *L. juvernica*, negative values to *L. sinapis*. The minimum error rate (probability for wrong identification) for each function is also given.

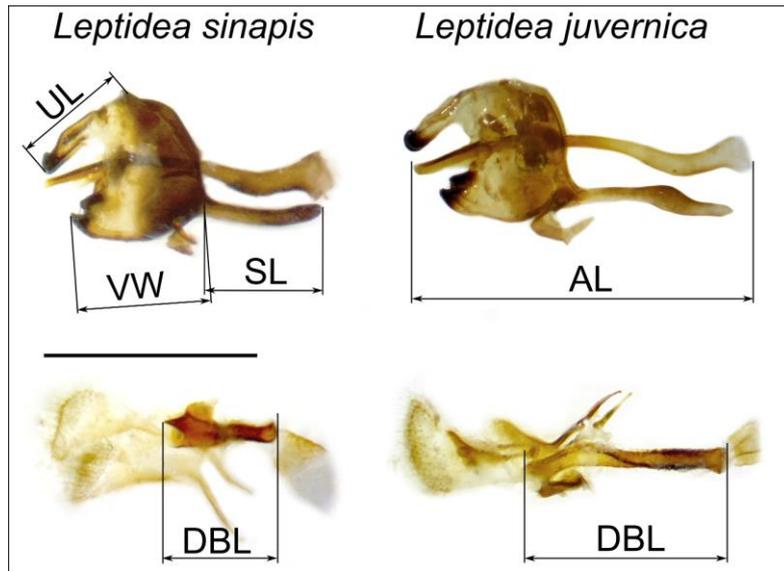| Variables	Discriminant function (dimensions in mm)	Method	Error rate (per 10 000)
AL, SL, VW, UL	AL + 0.262 SL – 0.618 UL – 0.845 VW – 0.891	LDA	4.95
AL, VW	**AL – 1.148 VW – 0.862**	**Scaling relations**	**7.90**
AL, VW	AL – 1.053 VW – 0.936	LDA	7.92
AL, SL	AL + 0.675 SL – 2.271	LDA	45.1
AL/VW, SL/VW	AL/VW + 0.545 SL/VW – 2.773	LDA	48.2
SL, VW	SL – 0.385 VW – 0.440	LDA	111

Figures

Figure 1. Male (top) and female (bottom) genitalia of *L. sinapis* (left) and *L. juvernica* (right). Measured dimensions are shown as follows: aedeagus length (AL), saccus length (SL), vinculum width (VW), and uncus length (UL) for males and ductus bursae length (DBL) for females. The scale bar is 1 mm.
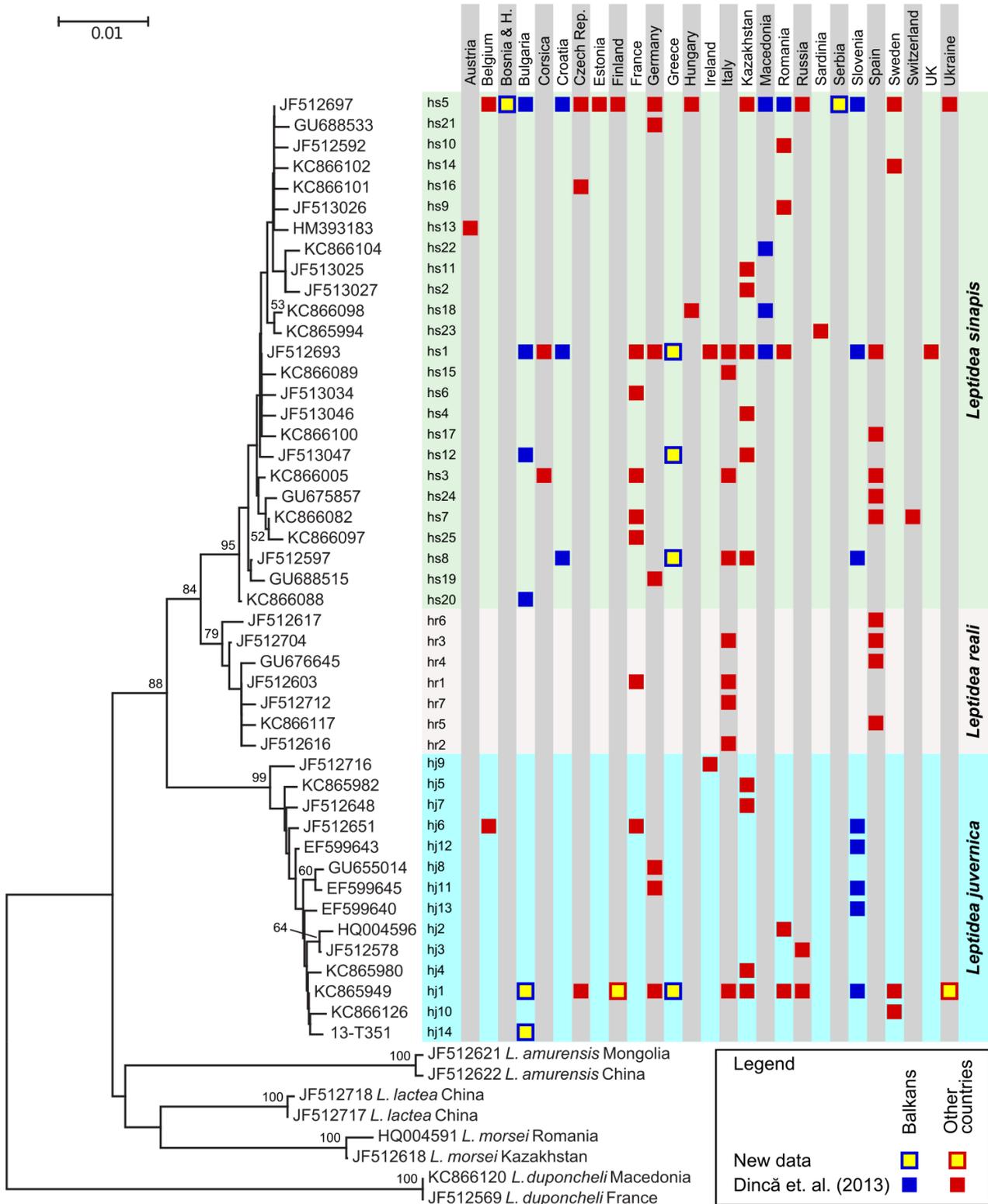
Figure 2. Neighbour-joining tree of mitochondrial *COI* haplotypes of *L. sinapis*, *L. reali*, and *L. juvernica*. *Leptidea amurensis*, *L. lactea*, *L. morsei*, and *L. duponcheli* were used as outgroup. The numbering and occurrence of the haplotypes is based on the combined datasets of this study and Dincă et. al. (2013). For the origin of all specimens and GenBank accession numbers, see Supplementary Table S1. The scale represents 0.01 substitutions per site. Bootstrap support values (1000 replicates) are shown next to the recovered nodes (only values higher than 50).
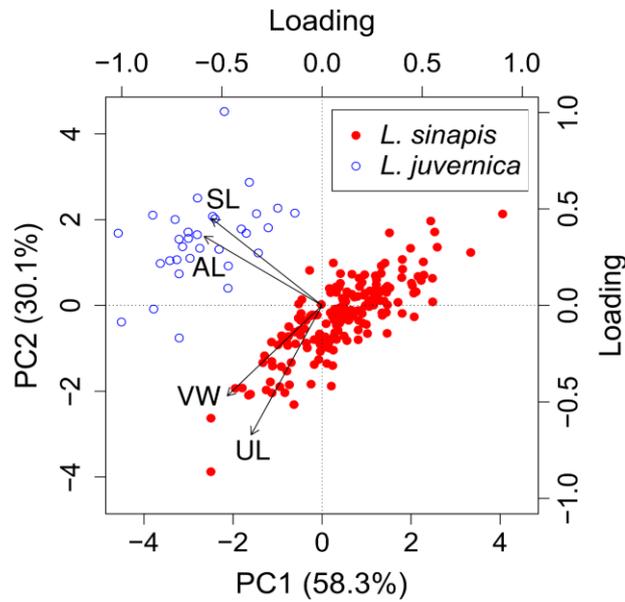
Figure 3. Results of principal components analysis of the morphometric data for male genitalia. The first two PCs are shown (bottom and left axes), as well as their loadings for the four measured variables (top and right axes). The identification based on a hierarchical clustering model (see text for details) is also shown for clarity but does not influence the analysis.
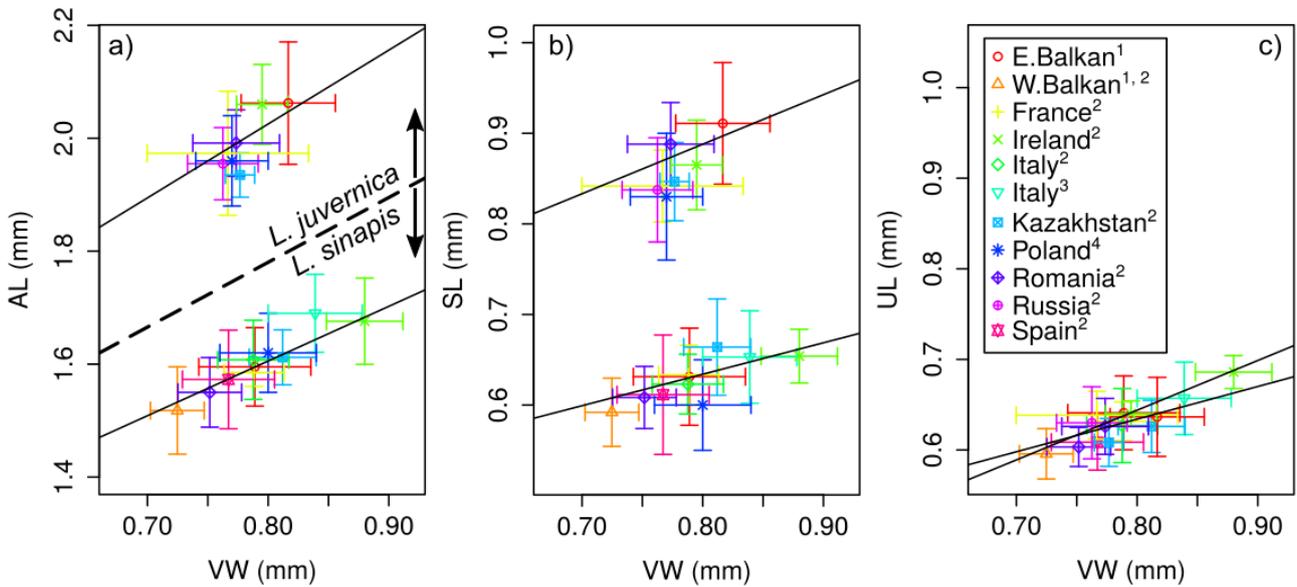

Figure 4. Scaling relations of genitalia dimensions with size (VW) by region: a) AL, b) SL, and c) UL. Linear fits to the combined Balkan and DLTV datasets are shown with straight lines. The error bars indicate standard deviations. Data are taken from the following sources (as indicated in the legend): 1: present study; 2: Dincă et al., 2011 (DLTV); 3: Fumi, 2008; 4: Sachanowicz, 2013. The discriminant function based on the AL scaling relation is illustrated with a dashed line and arrows in the left panel (a).
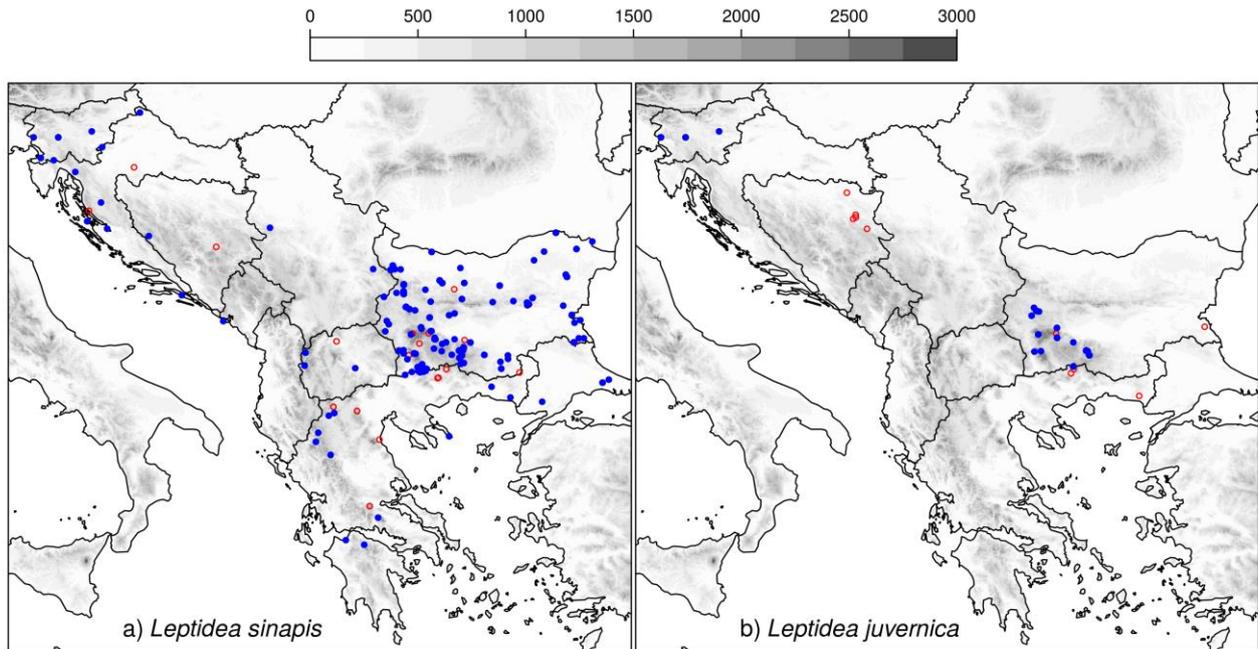
Figure 5. Distribution of a) *L. sinapis* and b) *L. juvernica* in the Balkan Peninsula. Data analysed in the present study are shown with filled blue circles, those from other sources (Kristal and Nässig, 1996; Rákosy, 1996; Mazel and Leestmans, 1999; Lelo, 2007; Pamperis, 2009; Coutsis, 2013) are shown with empty red circles.
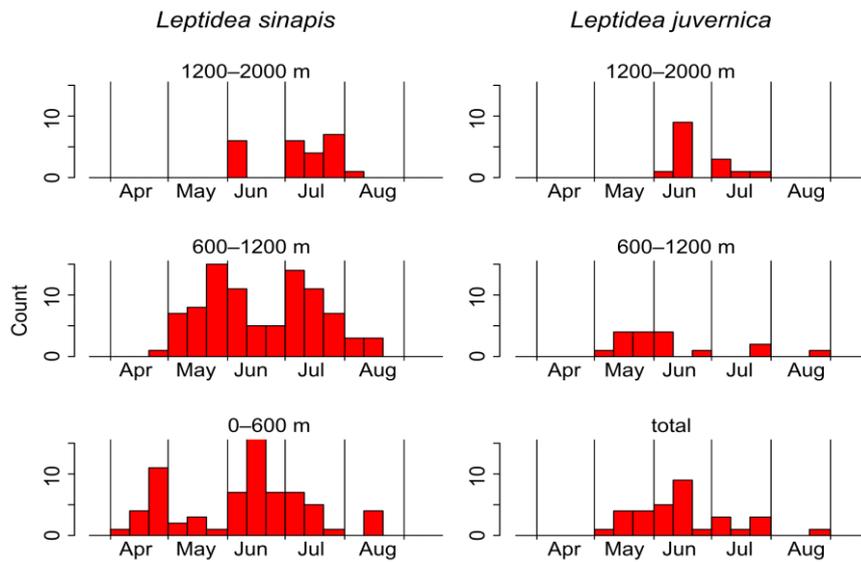

Figure 6. Flight period of *L. sinapis* (left) and *L. juvernica* (right). Only data from the eastern Balkans are included.

Supplementary information

Table S1. Specimens used for DNA analyses.

Table S2. Specimens used for morphological analyses.