MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 4267–4284 doi: 10.1111/mec.13756

Integrative analyses unveil speciation linked to host plant shift in *Spialia* butterflies

JUAN L. HERNÁNDEZ-ROLDÁN,*†[1] LEONARDO DAPPORTO,*‡[1] VLAD DINCĂ,*§ JUAN C. VICENTE,¶ EMILY A. HORNETT,** JINDRA ŠÍCHOVÁ,†† VLADIMIR A. LUKHTANOV,‡‡§§ GERARD TALAVERA*¶¶ and ROGER VILA*

*Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37, E-08003 Barcelona, Spain, †Departamento de Biología (Zoología), Facultad de Ciencias de la Universidad Autónoma de Madrid, C/ Darwin 2, E-28049 Madrid, Spain, ‡Department of Biology, University of Florence, Via Madonna del Piano 6, 50019 Sesto Fiorentino, FI, Italy, §Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, Canada N1G 2W1, ¶C/ Witerico, 9A – Bajo B, E-28025 Madrid, Spain, **Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK, ††Institute of Entomology, Biology Centre ASCR, 370 05 České Budějovice, Czech Republic, ‡‡Department of Karyosystematics, Zoological Institute of Russian Academy of Sciences, Universitetskaya nab. 1, 199034 St. Petersburg, Russia, §§Department of Entomology, St. Petersburg State University, Universitetskaya nab. 7/9, 199034 St. Petersburg, Russia, ¶¶Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, USA*

Abstract

Discovering cryptic species in well-studied areas and taxonomic groups can have profound implications in understanding eco-evolutionary processes and in nature conservation because such groups often involve research models and act as flagship taxa for nature management. In this study, we use an array of techniques to study the butterflies in the *Spialia sertorius* species group (Lepidoptera, Hesperiidae). The integration of genetic, chemical, cytogenetic, morphological, ecological and microbiological data indicates that the *sertorius* species complex includes at least five species that differentiated during the last three million years. As a result, we propose the restitution of the species status for two taxa often treated as subspecies, *Spialia ali* (Oberthür, 1881) stat. rest. and *Spialia therapne* (Rambur, 1832) stat. rest., and describe a new cryptic species *Spialia rosae* Hernández-Roldán, Dapporto, Dincă, Vicente & Vila sp. nov. *Spialia sertorius* (Hoffmannsegg, 1804) and *S. rosae* are sympatric and synmorphic, but show constant differences in mitochondrial DNA, chemical profiles and ecology, suggesting that *S. rosae* represents a case of ecological speciation involving larval host plant and altitudinal shift, and apparently associated with *Wolbachia* infection. This study exemplifies how a multidisciplinary approach can reveal elusive cases of hidden diversity.

Keywords: biogeography, butterflies, Lepidoptera, new species, phylogeny, speciation

Received 15 November 2015; revision received 25 June 2016; accepted 5 July 2016

Introduction

The discovery of new species can be achieved on two fronts: through study of unexplored areas, habitats or taxonomic groups, and by venturing deeper into known biodiversity to uncover cryptic species (two or more distinct species previously classified as a single one). The advent of DNA techniques, combined with improvements in phenotype analyses, has produced an exponential increase of descriptions of cryptic species over the past two decades (Bickford *et al.* 2007; Rowe *et al.* 2007; Smith *et al.* 2008; Zemlak *et al.* 2009; Puillandre *et al.* 2010). Indeed, considerable genetic divergences within traditionally recognized species, in

Correspondence: Roger Vila, Fax: +34 932211011; E-mail: roger.vila@csic.es

[1]These authors contributed equally to this work.

combination with other subtle differences, can highlight potential cryptic species that require further analysis (Vrijenhoek *et al.* 1994; Feulner *et al.* 2006; Grundt *et al.* 2006; Dincă *et al.* 2015).

The delimitation of species requires the application of one (or more) species concept(s). Although the biological concept of species has arguably been the most influential, sometimes it is difficult to directly test it and most insect species have been described on the basis of morphological traits, which by definition cannot delimit synmorphic taxa. As a consequence, the recognition of cryptic species must rely on a combination of multiple markers such as molecular, behavioural, morphological, cytological and ecological (Sanders *et al.* 2006; Beheregaray & Caccone 2007; Dincă *et al.* 2011a,b, 2013), often resulting in stronger evidence for the occurrence of different evolutionary pathways. For this reason, cryptic taxa are becoming models in the study of speciation processes (e.g. *Leptidea* spp., Dincă *et al.* 2011b; *Heliconius* spp., Heliconius Genome Consortium 2012).

Butterflies belong to one of the most diverse orders —the Lepidoptera—and yet are among the best-studied invertebrates. Virtually complete inventories of their diversity are supposed to exist for particular areas, such as Europe. Nevertheless, several cryptic butterfly species have been discovered in the last decade (tropical regions, Hebert *et al.* 2004; Burns *et al.* 2008; Giraldo *et al.* 2008; Asia, Lukhtanov *et al.* 2008, 2015; North America, Shiraiwa *et al.* 2014; Europe, Dincă *et al.* 2011a,b; Zinetti *et al.* 2013). Certain cases are remarkable for the subtle differences among species documented, which in some instances seem to be completely synmorphic and represent a challenge to our understanding of butterfly diversity (e.g. Dincă *et al.* 2011b).

Adult butterflies of the genus *Spialia* are small and often difficult to distinguish by external morphology (De Jong 1978). The genus is distributed in the Palaearctic and Africa, and the larvae are generally monophagous or oligophagous (Tolman & Lewington 2008). According to the revisions of the genus (De Jong 1974, 1978), the *sertorius* group includes two morphologically differentiated and parapatric species: *S. orbifer* (Hübner, [1823]) in eastern Europe and temperate Asia, and *S. sertorius* (Hoffmannsegg, 1804) in western Europe and north-western Africa. According to Lorković (1973), Fazekas (1986) and Hesselbarth *et al.* (1995), these species have contact zones in Slovenia, Croatia, the Carpathian Mountains and Austria. The lack of intermediates in these contact zones has been regarded as a confirmation of their specific status (Hesselbarth *et al.* 1995). Several authors consider *S. sertorius* to be a polytypic species represented by three subspecies: *S. s. sertorius* from continental Europe, *S. s. ali* (Oberthür, 1881) from North Africa, and *S. s. therapne* (Rambur, 1832) from Corsica and Sardinia (De Jong 1974, 1978; Tolman & Lewington 2008; Tshikolovets 2011), although a number of authors (e.g. Balletto *et al.* 2014; Kudrna *et al.* 2015) consider *S. therapne* as a distinct species. A recent study (Dincă *et al.* 2015) has documented the existence of two deeply diverged mitochondrial lineages of *S. sertorius* in the Iberian Peninsula, which suggests that this species may include still unrecognized cryptic diversity.

In this study, we combine molecular (mitochondrial and nuclear DNA markers), chemical (cuticular hydrocarbons), cytological (chromosome number and chromosomal location of major rDNA sites), morphological (geometric morphometry of male genitalia and wing shape and pattern), ecological (larval food plant, altitudinal preference) and microbiological (presence of *Wolbachia* infection) data to assess the occurrence of cryptic taxa in the *sertorius* group and to reconstruct the ecological and evolutionary mechanisms driving the diversification of the group.

Material and methods

Genetic analyses

Specimen sequencing. The mitochondrial marker cytochrome *c* oxidase subunit I (*COI*) was sequenced from 250 specimens, and 38 sequences were obtained from GenBank (Dincă *et al.* 2011c, 2015; Hausmann *et al.* 2011). The nuclear internal transcribed spacer 2 (*ITS2*) was sequenced from 58 specimens and the nuclear wingless (*Wg*) from 59 specimens that were selected to represent the main *COI* lineages (Table S1, Supporting information). These two markers are among the most variable nuclear markers used for butterflies. Instances of intra-individual nuclear variation were coded as ambiguities. In the case of *ITS2*, gaps were treated as missing data. All novel sequences obtained in this study have been deposited in GenBank (Table S1, Supporting information) and are also available in the data set DS-SPIALIA from the Barcode of Life Data System (http://www.boldsystems.org/). The DNA extraction, amplification, sequencing and alignment protocols are described in Appendix S1 (Supporting information).

Phylogenetic analyses and dating of phylogenetic events. Prior to tree inference, alignments were subdivided according to gene region and codon position and the optimal substitution models and partitioning schemes were selected using PARTITIONFINDER v1.1.1 (Lanfear *et al.* 2012) applying the Bayesian information criterion

(BIC). To test for gene tree discordance, a hierarchical likelihood ratio test was used as implemented in CONCATERPILLAR v1.7.4 (Leigh *et al.* 2008). Tree inferences for the CONCATERPILLAR analyses were carried out in RAXML v7.2.8 (Stamatakis 2006), assuming a single GTR substitution model for each sequence alignment. MRBAYES 3.2 (Ronquist *et al.* 2012) was used to infer Bayesian (BI) phylogenetic trees for *COI*, *ITS2* and *Wg* separately, as well as in a combined data set. Two independent runs with four chains of 10 million generations each (with a prerun burn-in of 25%) were inferred for each analysis. BEAST 1.8.0 (Drummond *et al.* 2012) was used on the combined matrix to estimate divergence times based on *COI* substitution rates. A strict molecular clock, a coalescent (constant size) tree prior and partitioning by marker were employed, and substitution models were chosen according to AIC values obtained in JMODELTEST 0.1 (Posada 2008). A normally distributed substitution prior was set between 0.0075 and 0.0115 (within the 95% confidence interval), corresponding to substitution rates widely used for invertebrates: a slower 1.5% uncorrected pairwise distance per million years (Quek *et al.* 2004) and a faster 2.3% (Brower 1994). The latter coincides with a recent estimate for the first part of *COI* (barcoding region) in beetles (Andújar *et al.* 2012). Parameters were estimated using two independent runs of 20 million generations, and convergence was checked using the program TRACER 1.6. To examine relationships among *COI* haplotypes, maximum parsimony haplotype networks were constructed using TCS 1.21 (Clement *et al.* 2000), with an 18-step connection limit. A more detailed description of the phylogenetic methods is provided in Appendix S1 (Supporting information).

Molecular species delimitations. To investigate putative species boundaries, we used the general mixed Yule-coalescent (GMYC; Pons *et al.* 2006; Fujisawa & Barraclough 2013) and Poisson tree processes (PTP; Zhang *et al.* 2013) models. The addition of Yule and coalescent signal into the data set (e.g. incorporating related outgroup taxa with some extent of population structure) has been demonstrated to be beneficial for the GMYC performance (Talavera *et al.* 2013). Thus, ultrametric phylogenetic trees including 1009 *COI* Pyrginae sequences (944 retrieved from GenBank plus the 288 *Spialia* COI sequences of this work, collapsed to unique haplotypes) were used as an input for both methods. Two input trees were tested: a Bayesian tree using BEAST as described above and a ML tree using RAXML and further normalizing branch lengths in PATHd8 (Britton *et al.* 2006). A single-threshold approach for GMYC was evaluated using the R package SPLITS, and ML and BI (bPTP) implementations were conducted for PTP on the web server http://species.h-its.org/ptp/ using default settings.

Chemical analyses

Gas chromatography and mass spectrometry. Cuticular hydrocarbons (CHCs) were extracted from forewings of 28 male and analysed using a Hewlett-Packard (Palo Alto, CA, USA) 5890A gas chromatograph coupled with an HP 5971A mass selective detector (details provided in Appendix S1, Supporting information). CHCs were identified on the basis of their mass spectra produced by electron impact ionization (70 eV). To reduce the bias due to the use of compositional data in multivariate analyses, we transformed the area following the method provided by Aitchison (1986): $Zij = \ln(Yi,j/g(Yj))$; where Yi,j is the area of peak i for individual j, g (Yj) is the geometric mean of the areas of all peaks for individual j, and Zi,j is the transformed area of peak i for individual j.

Cytological analyses

Chromosome number. Male gonads were stored in Carnoy fixative (ethanol and glacial acetic acid, 3:1) for 15–24 months at 4 °C and then stained with 2% acetic orcein for 7–15 days at 20 °C. Squash preparations were conducted as was previously described in Vershinina & Lukhtanov (2010). We counted the number of chromosomal bivalents in metaphase I and the number of chromosomes in metaphase II of male meiosis. In total, preparations from 35 males were analysed. Cell divisions were found to be relatively rare in *Spialia* during adult stage, and metaphase plates were observed only in 13 individuals.

Chromosomal location of major ribosomal DNA clusters. Spread chromosome preparations of mitotic and meiotic chromosomes of *S. orbifer*, *S. sertorius* and *S. rosae* were made from both female and male gonads of fourth-instar larvae as was previously described in Mediouni *et al.* (2004). We examined the number and distribution of ribosomal DNA (rDNA) clusters by fluorescent in situ hybridization (FISH) with 18S rDNA probe. In total, two larvae of *S. orbifer*, seven larvae of *S. sertorius* and five larvae of *S. rosae* were analysed. An unlabelled 1650-bp-long 18S rDNA probe was generated by PCR from the codling moth (*Cydia pomonella*) genomic DNA (gDNA) extracted from adults by standard phenol–chloroform extraction as described in Fuková *et al.* (2005). The probe was labelled with biotin-16-dUTP (Roche Diagnostics GmbH, Mannheim, Germany) by nick translation using a Nick Translation Kit

(Abbott Molecular Inc., Des Plaines, IL, USA). FISH with the 18S rDNA and image processing were carried out as described in Fuková *et al.* (2005).

Morphological analyses

Geometric morphometry. A combination of landmarks and sliding semi-landmarks (Bookstein 1991) was applied to the outline of the left valva of the male genitalia and to the vein junctions and the outline of the white spots of the underside of the hindwing. Genitalic structures, valvae in particular, display great variability among butterfly taxa, and the underside of the hindwings in *Spialia* presents a complex pattern that is used as one of the main features in separating supposed species (De Jong 1974, 1978). Points that could be precisely identified were considered as landmarks, whereas other points were allowed to slide along the outline trajectory (sliding semi-landmarks). We identified eight landmarks and 24 semi-landmarks on the valva and 27 landmarks and eight semi-landmarks on the hindwings (Fig. S1, Supporting information). A generalized procrustes analysis (GPA) was applied to the landmark data to remove non-shape variation and to superimpose the objects in a common coordinate system. Partial warps were calculated using the shape residuals from GPA. Applying principal component analyses (PCA) to partial warps, relative warps (PCs) were obtained and used as variables in subsequent analyses (Bookstein 1991).

Analysis of configurations for genetic and phenotypic markers. As a first step, we performed a series of exploratory analyses to understand the main patterns of variation in the genetic and phenotypic markers and their degree of concordance. For this reason, the dissimilarity matrices for *COI*, *ITS2* and *Wg* were projected in two dimensions by principal coordinate analysis (PCoA) using the 'cmdscale' R function. Bidimensional representations were also provided for the three phenotypic characters (morphology of wings and genitalia and CHCs). For genitalia and wing shape, the first two relative warps (obtained with a PCA from partial warps) were used. For CHCs, we applied PCA to the transformed GC peak areas. To facilitate a direct comparison of their patterns, we eliminated the effect of location and rotation among bidimensional representations with procrustes analyses using the *COI* configuration as reference. Because different markers were represented by different sets of specimens, we used the 'recluster.procrustes' function of the recluster R package, which maximizes similarities among configurations on the basis of partially overlapping data sets (Dapporto *et al.* 2014a). The bidimensional configurations for specimens were projected in RGB colour space using the same package (Dapporto *et al.* 2014b). Specimens belonging to the same grid square of 2° for latitude and longitude were grouped, and their individual RGB colours were plotted on a map in pie charts.

In a second step, we tested the existence of a signature of diversification in CHC composition and genitalic and wing shape among the supposed set of species identified by different markers (genetic, microbiological, ecological). We performed a series of partial least square discriminant analyses (PLSDA) with shape variables (PCs) and transformed GC peak areas as variables and the hypothesis for species attribution as grouping variable. As relative warps can be particularly numerous (2 × number of landmarks-4), as well as the compounds existing on the cuticle, overfitting had to be avoided. We thus applied a sparse PLSDA (Lê Cao *et al.* 2011) only selecting the most influential five variables for each resulting component. To evaluate the degree of diversification as a percentage of cases that can be blindly attributed to their group, we applied a jackknife (leave-one-out) algorithm and individually classified each specimen. These analyses were carried out with the 'splsda' and 'predict' functions in the mixOmics R package (Lê Cao *et al.* 2011).

Microbiological analyses

Presence and identification of Wolbachia strains. A total of 102 specimens were surveyed for the presence of the heritable bacterial endosymbiont *Wolbachia* (details provided in Appendix S1, Supporting information). To determine the strain(s) of *Wolbachia*, 10 specimens representative of the three infected taxa (*S. therapne*, *S. orbifer* and *S. rosae*), from a variety of localities and with diverse mitochondrial (*COI*) haplotypes, were additionally analysed using primer pairs that amplified four further loci (*hcpA*, *gatB*, *ftsZ* and *fbpA*). These, together with *coxA*, make up the multilocus sequence typing (MLST) system for *Wolbachia* (Baldo *et al.* 2006). For each specimen, PCR product for each of the five MLST genes, as well as for the hypervariable *wsp* gene, was sequenced using the Sanger technology. Sequences were then compared to existing records using the *Wolbachia* MLST Database (pubmlst.org/wolbachia/) in order to identify the sequence type for each gene locus, and then to build an allelic profile for each strain. The concatenated allele sequences for each different strain of *Wolbachia* found in *Spialia* ($n = 4$) were then used to create a phylogeny alongside concatenated MLST sequences representing a variety of *Wolbachia* strains infecting Lepidoptera and other insects, retrieved from the *Wolbachia* MLST database. Sequences were aligned and maximum-likelihood phylogenetic trees were

constructed in MEGA6.06 (http://www.megasoftware.net/). Tree support was evaluated by bootstrapping with 1000 replications. The tree was rooted using the *Wolbachia* strain infecting the nematode *Brugia malayi*. *Wolbachia* supergroup for each clade was identified and indicated on the tree (supergroups A, B, D and F).

Ecological analyses

Larval host plants and altitudinal specialization. Plants of *Sanguisorba* spp. and *Rosa* spp. (all species found in each locality) were inspected in search of both eggs and larval refugia. In addition, female *Spialia* were followed in the field to observe oviposition behaviour. The plants were determined by plant taxonomists (see Acknowledgements) based on morphology, and the butterfly immatures were sequenced at least for the marker *COI*. To test whether the putative species display different altitudinal specialization, we identified all the localities where at least one specimen was recorded. We compared the altitudes for all the sites of the different species by applying a Monte Carlo Kruskal–Wallis test based on 10 000 iterations. We then applied a Tukey post hoc test among the five taxa using the PCMCR R package.

Results

Genetic analyses

According to the Bayesian phylogenetic tree inferred from *COI* sequences (Table 1, Fig. S2, Supporting information), the taxa *therapne* from Corsica and Sardinia and *ali* from North Africa were recovered as deeply diverged taxa with high support that, together, were sister to the other taxa. All species delimitation analyses based on *COI* sequences (Fig. S3, Supporting information) recovered *S. therapne*, and nominotypical *S. sertorius* as putative species, but split *S. ali* into two entities and *S. orbifer* into three to four entities (all supported in the Bayesian *COI* tree). In terms of spatial genetic structure, the Corsican specimens of the taxon *therapne* represented a distinct haplotype (t2) from the Sardinian ones (Fig. S4, Supporting information). The two main lineages within *S. ali* displayed a minimum uncorrected p-distance of 3%, were both detected in Morocco, Algeria and Tunisia, and were found in sympatry in north-western Tunisia (haplotypes a5 and a6) (Fig. S4, Supporting information). The *COI* haplotype network for specimens of *S. orbifer* and nominotypical *S. sertorius* and (Fig. 1) presented five loops, which were broken according to frequency and geographical criteria (Excoffier & Langaney 1989). *Spialia sertorius* displayed a widespread haplotype (S1) and endemic haplotype clades in the Iberian and Italian Peninsulas (Fig. 1). The four main lineages of *S. orbifer* were reflected as distinct clades in the *COI* haplotype network with the following distributions: one lineage was detected in Sicily, Romania and eastern Kazakhstan; another one was found in the Balkans, Turkey and Armenia, a third one was found only in eastern Kazakhstan, and the fourth was confined to the Iberian Peninsula (from here on we refer to it as *rosae*) (Fig. 1). The *rosae* clade displayed a 1% minimum *COI* p-distance and was geographically far from the other *S. orbifer* lineages, but it was sympatric with several Iberian populations of *S. sertorius* (Fig. 1). The *S. orbifer* lineage detected exclusively in eastern Kazakhstan was the most diverged and displayed a minimum genetic p-distance of 1.8% to the nearest conspecific clade. This lineage was separated by only 100 km from the nearest conspecific lineage that also occurred in eastern Kazakhstan (Fig. 1).

Table 1 Summary of differences among species of the *Spialia sertorius* species group

Species	*COI*	*ITS2*	*Wg*	CHCs%	Wings%	Genitalia%	Host plant	Karyotype	*Wolbachia*%
S. ali	1	0.96	1	100.0	100.0	90.9	*Sanguisorba* spp.	$n = 31$	0
S. therapne	1	0.99	U/M	75.0	100.0	87.5	*Sanguisorba* spp.	$n = 31$	100
S. orbifer	N/M	U/M	N/M	100.0	93.1	38.9	*Sanguisorba* spp.	$n = 31$	56
S. sertorius	1	N/M	N/M	83.3	24.0	22.5	*Sanguisorba* spp.	$n = 31$	0
S. rosae	0.98	N/M	N/M	100.0	46.4	50.0	*Rosa* spp.	$n = 31$	100

Red, yellow and blue indicate strong, moderate or no differentiation, respectively. For the genetic markers (*COI*, *ITS2* and *Wg*), posterior probabilities (p.p.) of the nodes defining monophyly of each species in the Bayesian single marker trees are indicated. Blue, monophyletic with p.p. =>95; yellow, unsupported monophyly (U/M) with p.p. <95; blue, not monophyletic (N/M). Percentage of correct attribution of specimens to the five entities as obtained by jackknife partial least square discriminant analysis for cuticular hydrocarbons (CHCs), wing shape (Wings) and genitalia shape (Genitalia), (blue, less than 75% identification power; yellow, between 75 and 90% of identification power; red, identification power higher than 90%). Host plant: host plants as indicated by literature and by our field data. Karyotype: haploid chromosome number based on our results. *Wolbachia*: percentage of the specimens of each species that were infected in our analyses.

Similarly to *COI*, the Bayesian analysis of *ITS2* sequences (Table 1, Fig. S5, Supporting information) supported the monophyly of *S. therapne* and *S. ali* and recovered them as a clade sister to the other taxa. However, unlike *COI*, the nominotypical *S. sertorius* was not monophyletic and mixed with specimens of the *rosae* lineage. Specimens corresponding to the three other *S. orbifer* lineages formed a clade, albeit with low support, and with a notable differentiation in specimens from Turkey and Armenia, which formed well-supported clades. The Bayesian phylogenetic analysis of *Wg* sequences (Table 1, Fig. S6, Supporting information) also recovered the taxa *therapne* and *ali* as monophyletic, but not the rest of the taxa, and deeper relationships were not well resolved. Topological incongruence between individual gene trees was analysed through a hierarchical likelihood ratio test with CONCATERPILLAR. The test suggested discordance between *COI-Wg* and *ITS2*. We thus explored and compared trees combining *COI-Wg* and all three markers. The combined analysis of *COI* and *Wg* sequences (Fig. S7, Supporting information) produced results virtually identical to those based exclusively on *COI*, with considerably improved supports for the monophyly of the whole *sertorius* species group, as well as for the sister relationship between *S. sertorius* and *S. orbifer+rosae*. An analysis using all three markers (*COI*, *ITS2* and *Wg*) (Fig. 2) produced results that were in line with those based on *COI*, as well as on *COI* and *Wg* combined. The exception was the exact position of *S. rosae* sp. nov., which was recovered as sister to a monophyletic *S. orbifer*. Sorting effects or, alternatively, introgression could explain the differential signal obtained for this taxon using different markers. Species delimitation by GMYC, PTP and bPTP models for the *Spialia* tree obtained with the combined data set (Fig. S8, Supporting information) was similar to those based on *COI* (Fig. S3, Supporting information): they generally supported the species status for the five species we propose in Fig. 2 and suggested that further cryptic entities may exist within *S. ali* (two entities) and *S. orbifer* (between 3 and 5 entities).

The Bayesian chronogram based on *COI*, *Wg* and *ITS2* sequences (Fig. S9, Supporting information) suggested that the *sertorius* species group diversified roughly during the last three million years, with the split between *S. therapne* and *S. ali* taking place ca. two million years ago (mya), and diversification within the group formed by *S. sertorius* and *S. orbifer* starting ca. 1.5 mya. It is worth noting that age estimates based on published substitution rates for other taxa necessarily involve substantial uncertainty.

Chemical analyses

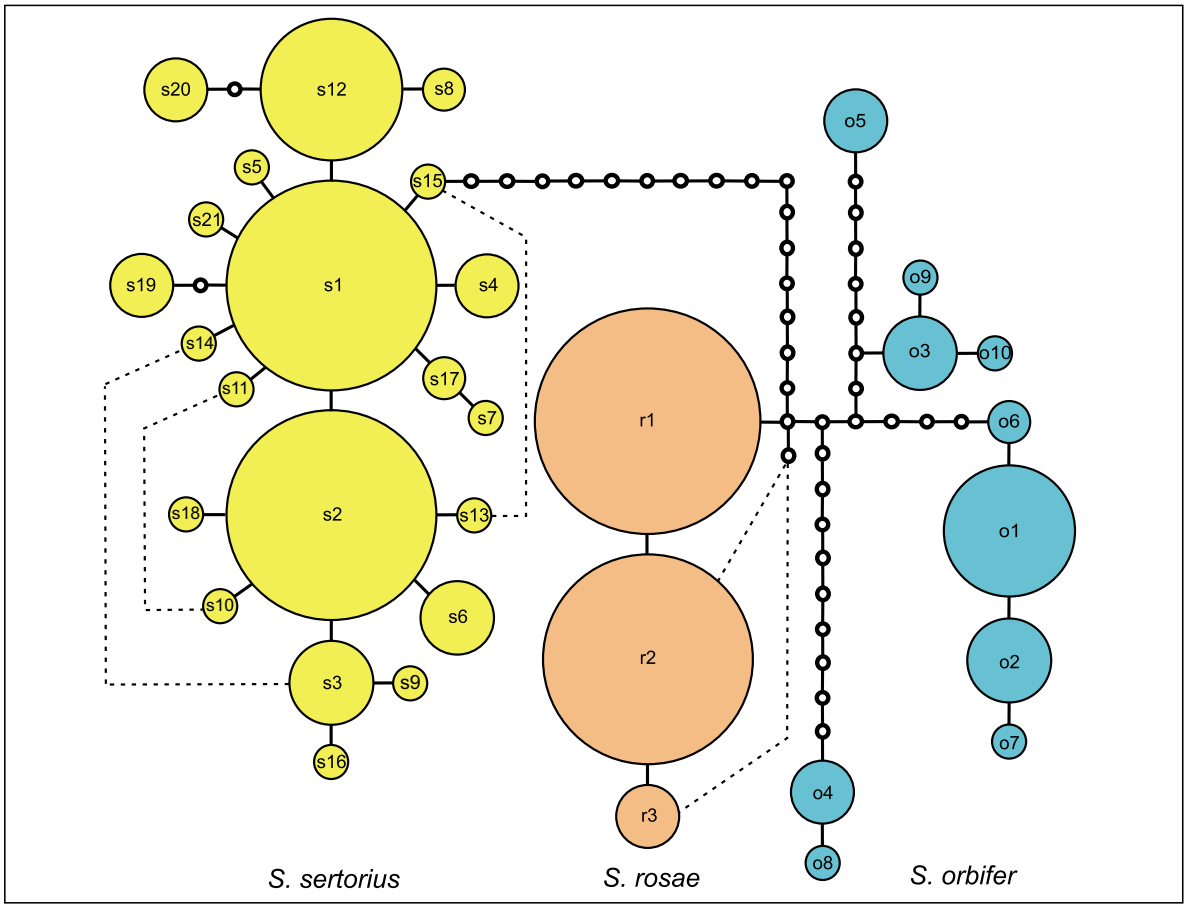
We found a total of 23 CHCs on the wings of the 28 specimens analysed. All these chemicals represent linear alkanes from C23 to C31 and C33 (10 compounds), seven methyl-branched alkanes, two dimethyl-branched alkanes, three alkenes and one alcohol (see Appendix S1, Supporting information for details). A scatterplot of a PLSDA showed that *S. ali* + *S. therapne*, *S. sertorius*, *S. orbifer* and *S. rosae* form four clusters (Fig. 3), only slightly overlapping. *Spialia ali* and *S. therapne* were characterized by profiles dominated by linear compounds (Fig. 4), except for one unusual *S. therapne* specimen. By contrast, unlike all other species, *S. rosae* showed a striking pattern characterized by the abundance of C27:1, C29:1, 7methylC27, central-methylC26 and 11,Y dimethylC25. Moreover, it showed the lowest abundances (near to zero) of C25:1, 2methylC26 and 2methylC28, compounds that are well represented in *S. sertorius* and *S. orbifer* (Fig. 3, Fig. S10, Supporting information). The jackknife partial least square discriminant analysis attributed virtually all the samples (except one *S. therapne*) to their original group (96.4% of correct assignments) (Table 1). This result demonstrates that there is a combination of compounds that allows an almost complete discrimination among the five supposed taxa.

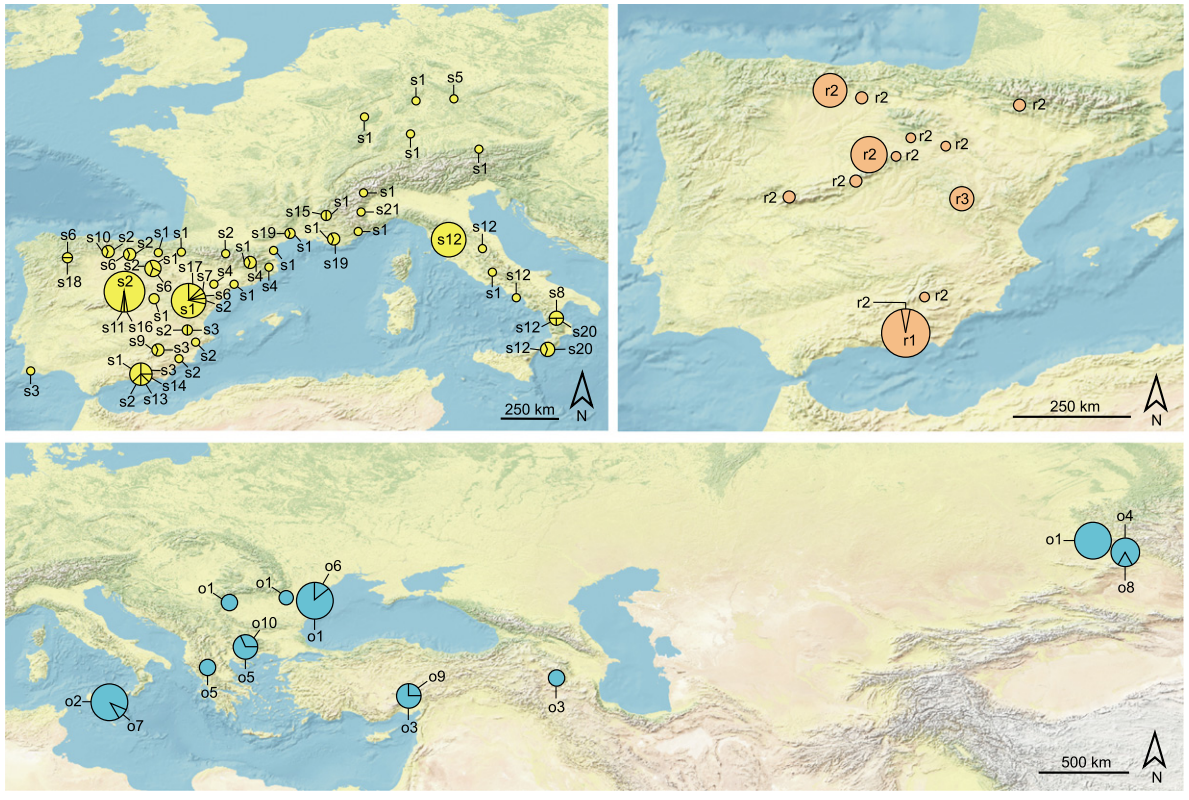
Cytological analyses

Chromosome number. In all the studied individuals of the taxa *S. rosae*, *S. orbifer* and *S. therapne,* the same karyotypes with $n = 31$ were found (Table S2, Supporting information). In metaphase I (MI), the cells possessed 31 bivalents, while in metaphase II (MII) the cells had 31 chromosomes (Fig. S11, Supporting information). The bivalents and chromosomes were not strongly differentiated with respect to their size, and the largest elements (bivalents and chromosomes) were only 2–2.5 times larger than the smallest ones. The size of all 31 bivalents in MI stages and of all 31 chromosomes in MII stages decreased more or less linearly.

Chromosomal location of major rDNA. FISH with 18S ribosomal DNA (rDNA) probe did not reveal any difference

Fig. 1 Mitochondrial *COI* haplotype networks and their geographical distribution for *S. sertorius* (yellow), *S. rosae* sp. nov. (pink) and *S. orbifer* (blue). The size of the circles is proportional to the number of samples displaying the haplotype. Connections creating loops that are less probable according to frequency and geographical criteria are indicated in dashed lines. Haplotype codes match those in Table S1 (Supporting information).





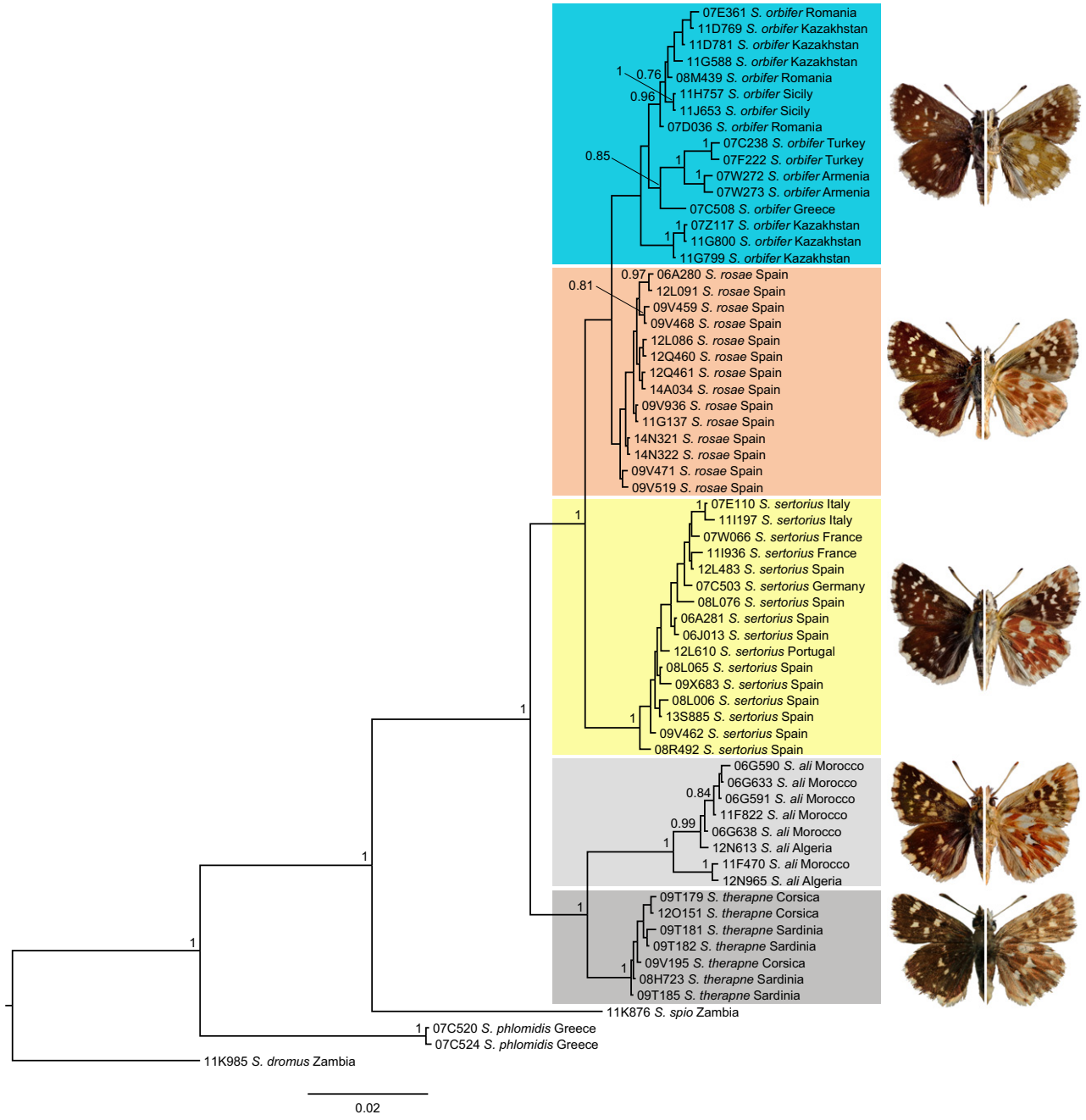


Fig. 2 Bayesian phylogeny for the *Spialia sertorius* species group based on the combined analysis of *COI*, *ITS2* and *Wg*. Species proposed in this study are highlighted by coloured blocks. Bayesian posterior probabilities higher than 0.7 are shown next to the recovered branches.

in the number and location of major rDNA clusters in all studied larvae of *S. orbifer* and *S. rosae* In both taxa, the rDNA probe localized at the end of a single bivalent in the pachytene stage (Fig. S12a, c, Supporting information), and at the ends of two mitotic metaphase chromosomes (Fig. S12b, d, Supporting information). These results clearly indicate the presence of a single pair of chromosomes, each carrying a cluster of rRNA genes forming a nucleolar organizer region (NOR). In larvae of *S. sertorius*, we found intraspecific differences in the number of rDNA sites. In pachytene nuclei of five larvae, we observed two terminal clusters, that is one large and one small (Fig. S12e, Supporting information), but only three hybridization signals in mitotic chromosomes (Fig. S12f, Supporting information). In mitotic metaphase complements of other two *S. sertorius* larvae, we

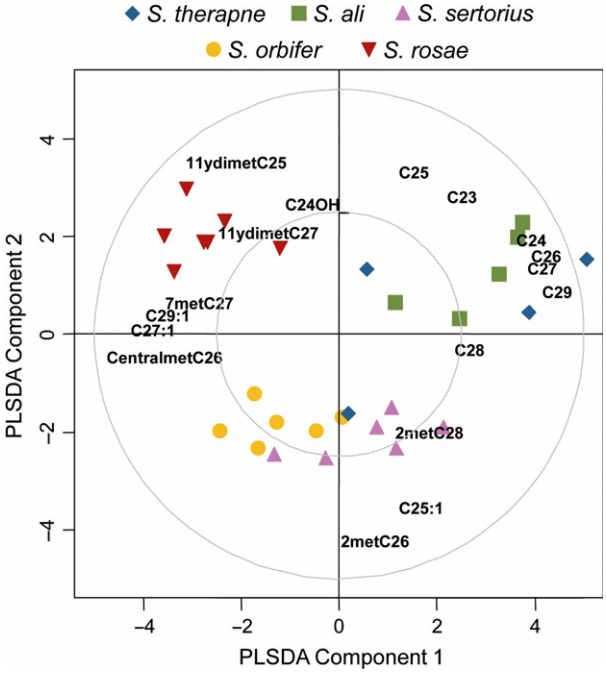


Fig. 3 Partial least square discriminant analysis of cuticular hydrocarbon composition for the five species. The pattern of both chemical compounds (variables) and individual (cases) is shown in the biplot. The two concentric circles represent variable loadings of 0.5 and 1. C23–C29 represent linear alkanes; C25:1-C29:1 alkenes; 2metC26 and 2metC28 are C26 and C28 methylated in position 2; 7metC27 is a C27 methylated in position 7; CentralmetC26 is a mixture of compounds with central methyl substituents; 11ydimetC27 and 11ydimetC29 are C27 and C29 alkanes dimethylated in position 11 and in another unknown position; C24OH is an alcohol.

found only two terminal rDNA sites (Fig. S12g, Supporting information). The rDNA heterozygosity suggests the presence of chromosomal rearrangements in the karyotype of *S. sertorius*. Moreover, counts of mitotic metaphase chromosomes on these preparations confirmed the haploid chromosome number of $n = 31$ in all analysed *S. sertorius* larvae (Fig. S12b, d, f, g, Supporting information).

Morphological analyses

Wing pattern. The analysis of the hindwing resulted in 66 relative warps (PCs). The first two shape components explained together only a reduced fraction of the variance (25.68%), and the unconstrained PCoA revealed a very slight separation among the five supposed taxa and a poor geographical structure (Fig. 4), thus confirming a high morphological similarity within the *sertorius* group. However, the jackknife PLSDA revealed that a large fraction of *S. ali*, *S. therapne* and *S. orbifer* specimens can be blindly attributed to their species on the basis of a combination of wing shape variables. Only, *S. sertorius* and *S. rosae* revealed to be indistinguishable on the basis of hindwing shape and pattern (Table 1, Figs S13 and S14, Supporting information).

Male genitalia. The analysis of the genitalia pattern resulted in 62 relative warps (PCs). The first two components explained 42.60% of the variance. As it occurred for wing shape, the first two components of genitalia shape revealed neither differential patterns among the five species, nor spatial structure in the PCoA (Fig. 4). However, the jackknife PLSDA showed that *S. therapne* and, even more, *S. ali* specimens can be distinguished on the basis of a complex combination of shape variables, while *S. sertorius*, *S. orbifer* and *S. rosae* were indistinguishable (Table 1, Figs S15 and S16, Supporting information).

Microbiological analyses

Presence and identification of strains of Wolbachia. We tested 102 specimens for the presence of the endosymbiont *Wolbachia* (Table S1, Supporting information). Overall 44% (45/102) were infected with *Wolbachia*; however, infection rate differed widely among *Spialia* species. While every specimen of two species (*S. rosae* and *S. therapne*) was infected, a third species (*S. orbifer*) was polymorphic for infection (i.e. some specimens were infected and some were not), and the remaining two species (*S. sertorius* and *S. ali*) showed no evidence of infection. Indeed, *S. orbifer* infection rate varies with the location of the population; for example, 100% of samples from Kazakhstan were infected, while no Italian sample (isolated population in Sicily) showed the presence of *Wolbachia*. It should be noted that in some cases (particularly Bulgaria and Greece) the sample size was too small to conclude that *Wolbachia* is absent from these populations (Table S3, Supporting information).

Ten specimens were then selected for multilocus sequence typing (MLST) to identify which *Wolbachia* strain(s) were present in *Spialia*. These included two *S. therapne* (Sardinia and Corsica), six *S. orbifer* (Romania, Armenia, Turkey and three from Kazakhstan), and two *S. rosae* (northern and southern Spain). Using the MLST system, four *Wolbachia* strains were detected across the samples, one in both *S. therapne* and one of the *S. rosae* specimens (strain 374), a second in the second *S. rosae* (strain 160), a third in *orbifer* from Turkey and Armenia (strain 296), and a fourth from the remaining *orbifer* specimens collected in Kazakhstan and Romania (strain 300). However, when combined with *wsp* sequence data, the fourth strain (300) of *Wolbachia* was further split into two variants. One specimen of *S. rosae* and one of *S. therapne* were apparently infected

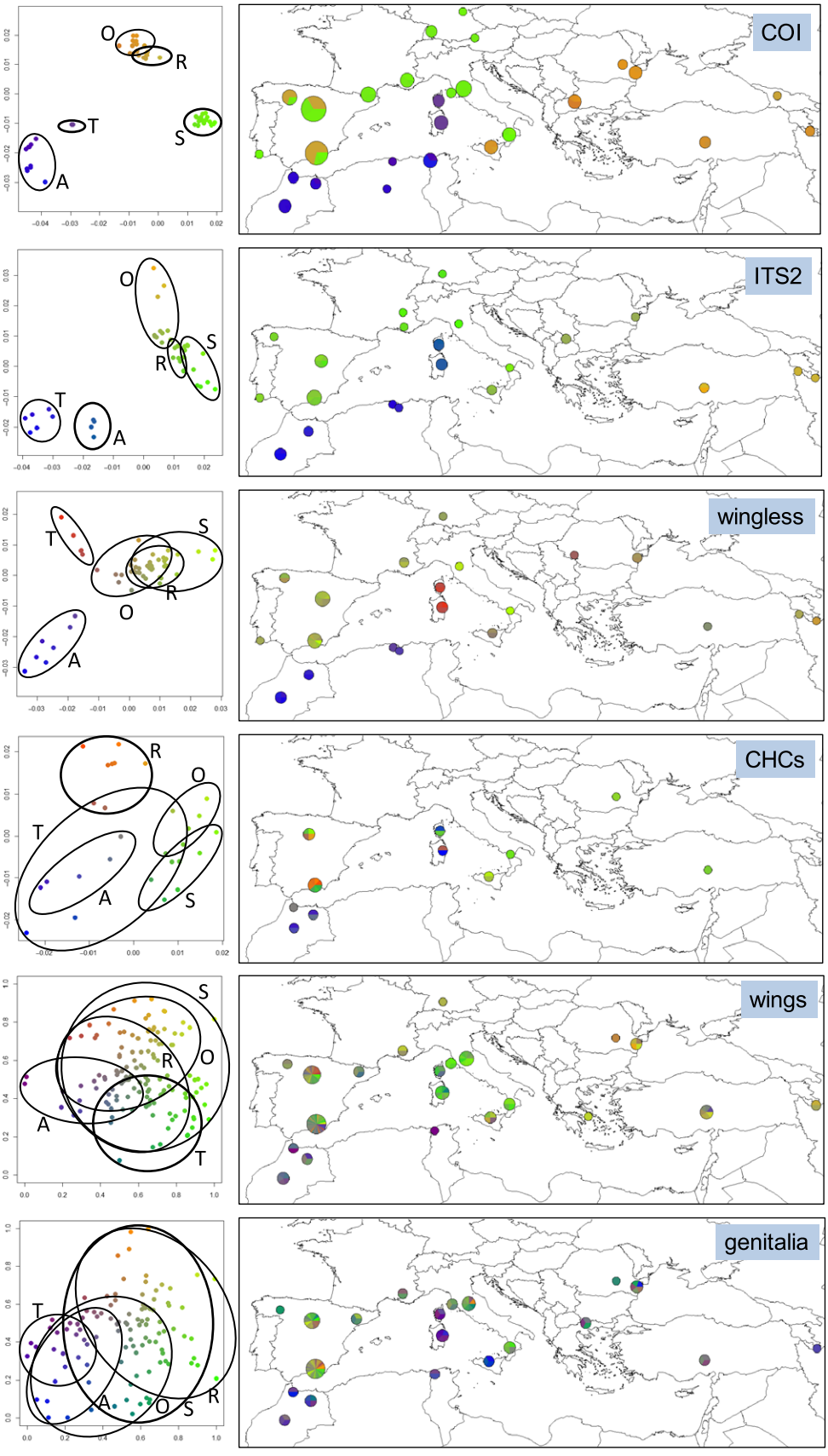


Fig. 4 Results of unconstrained analyses for *COI*, *ITS2* and *Wg* genetic markers, for cuticular hydrocarbon composition (CHCs), and for wings and genitalia morphology. On the left, representations of principal coordinate analyses based on dissimilarity matrices for each marker. The specimens have been projected in RGB colour space, and the resulting colours were plotted in pie charts grouping specimens from the same island or continent within 5 degrees latitude–longitude squares (maps on the right). A = *S. ali* stat rest., O = *S. orbifer*, R = *S. rosae* sp. nov., S = *S. sertorius*, T = *S. therapne* stat. rest.

simultaneously by two different strains of *Wolbachia*. Phylogenetic analyses revealed that two of the strains detected in *Spialia* grouped with *Wolbachia* supergroup A, one with B, while the fourth was grouped with supergroup F (Fig. S17, Supporting information).

Ecological analyses

Larval host plants. All the taxa in the studied species group have been described as monophagous on *Sanguisorba* spp. (especially *Sanguisorba minor*), although sporadically they have been reported as also feeding on species of the genus *Potentilla* and *Rubus* (Rosaceae) (e.g. Tolman & Lewington 2008; Tshikolovets 2011). We obtained oviposited eggs and larvae within refugia on *Sanguisorba minor* found in the wild for the species *S. sertorius*, *S. ali* and *S. orbifer* (Table S1, Supporting information). A notable exception was the case of *S. rosae*. Several females of this species were observed ovipositing, and multiple larvae were collected within refugia, on various species of *Rosa* spp. (Tables S1 and S4, Supporting information). All the specimens collected on *Rosa* spp. were determined as *S. rosae* based on the *COI* mitochondrial marker, while all those from the

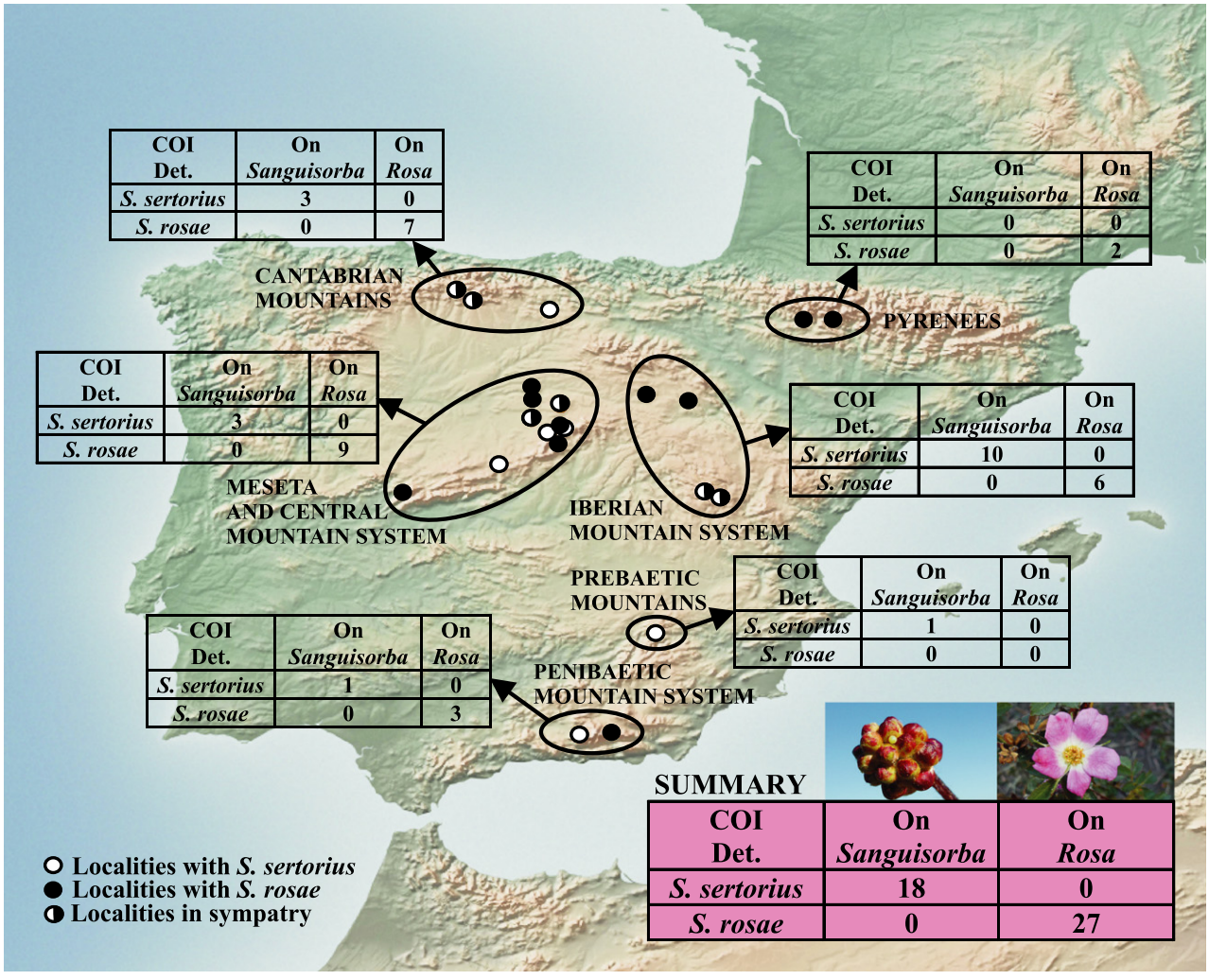


Fig. 5 Summary of *Spialia sertorius* and *Spialia rosae* sp. nov. *COI* and host plant results in Iberia. The two characters invariably match at specimen level in sympatry, which demonstrates the existence of two ecologically differentiated species.

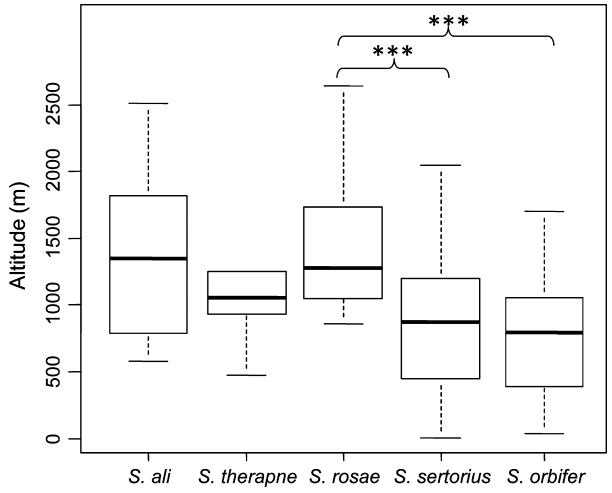


Fig. 6 Boxplots reporting median values and interquartile ranges of the altitude for the collection localities of the five species within the *Spialia sertorius* species group. Asterisks mark the *P* value obtained by Monte Carlo Mann–Whitney comparisons (***$P < 0.001$, we did not consider values between 0.05 and 0.01 as significant after Bonferroni correction).

Iberian Peninsula collected on *Sanguisorba minor* were *S. sertorius*, even if they were in many instances collected in sympatry (Fig. 5, Table S4, Supporting information).

Altitudinal specialization. The five putative species showed an overall significantly different altitudinal pattern (Monte Carlo Kruskal–Wallis, $P < 0.001$). However, when performing pairwise comparisons only *S. rosae* showed a significantly higher altitudinal range than the closely related *S. sertorius* and *S. orbifer* (Fig. 6).

Discussion

Speciation in the genus Spialia

Evolutionary events and ecological mechanisms generating and maintaining diversity in the *Spialia sertorius* species group appear to be complex. However, the high-resolution spatial assessment, the analysis of immature stages directly collected on host plants, and the combination of several molecular, phenotypic and microbial data helped us to better understand the evolutionary history of this species group and document the existence of at least five distinct species.

Allopatric species. In the case of allopatric taxa, it is virtually impossible to study gene flow and barriers to reproduction, except when there is or has recently been some degree of dispersal among them. Thus, although we conceptually rely on the biological species concept, it can hardly be tested directly in allopatry. Consequently, we consider allopatric taxa as species when they display a comparable or higher level of morphological/molecular differentiation than among sympatric or parapatric species and their lumping would result in well-supported nonmonophyletic assemblages.

Although some authors (Balletto *et al.* 2014; Kudrna *et al.* 2015) and some resources on Lepidoptera (IUCN Red List and Fauna Europaea) consider *S. therapne* as a species, this taxon (endemic to Corsica and Sardinia) as well as the taxon *ali* (North Africa) has been generally treated as subspecies of *S. sertorius* (e.g. De Jong 1974, 1978; Tolman & Lewington 2008; Tshikolovets 2011). Our genetic analyses with all markers (individually and combined) recovered the allopatric *S. ali* and *S. therapne* as strongly divergent clades that were sister to the other species (Fig. 2). If they are not considered species, their classification as subspecies of *S. sertorius* would render this species as a deeply paraphyletic entity with respect to the widely accepted *S. orbifer*. Thus, we propose a specific status for both *Spialia ali* (Oberthür, 1881) stat. rest. and *Spialia therapne* (Rambur, 1832) stat. rest. In addition, all species delimitation methods supported these two taxa as species, but they further suggested the potential existence of cryptic taxa within *S. ali* (two entities). The two lineages of *S. ali* apparently diverged during the last million of years in North Africa. They are sympatric, were collected on the same host plant species, and may represent either inter- or intraspecific variability, but further study is needed to clarify their status. Although we obtained a perfect identification of specimens based on wing pattern and a good identification based on genitalia (Table 1), the overall pattern for morphology reveals only a slight diversification (Fig. 4). Therefore, the identification of specimens relies on minor morphological features as the first principal components did not reveal any clustering of specimens. The results for CHC composition reinforce the hypothesis that *S. ali* and *S. therapne* are different taxa. Actually, based on a combination of compounds, all the specimens except one can be attributed to their species group. A recent study on the genus *Pyrgus*, which belongs to the same subfamily as *Spialia* (Pyrginae), showed that chemical divergence among populations tends to be rather weak, but it rapidly rises and attains a plateau when speciation takes place, as expected if they represent species-specific recognition cues (Hernández-Roldán *et al.* 2014).

It is worth noting that all *S. therapne* specimens analysed were infected by the bacterial endosymbiont *Wolbachia*, but no infection was detected in specimens of *S. ali*. The divergence between *S. ali* and *S. therapne* is dated at ca. 2 My, supposedly when a dispersal event between Africa and the islands of Sardinia and Corsica took place. These islands are considered among the most important areas of endemism in the

Mediterranean, probably because of the long isolation and of the suitable climate for many species during the Pleistocene glaciations (Schmitt 2007). The identification of the taxa from North Africa as the closest relatives to the taxa occurring on Sardinia and Corsica compared to the closer Italian mainland confirms a well-known paradigm for butterflies found in most of the Satyrinae (Kodandaramaiah & Wahlberg 2009; Dapporto *et al.* 2012) and in some Lycaenidae (Vodă *et al.* 2015).

Parapatric and sympatric species. When dealing with taxa that coexist in some parts of their ranges, it is possible to actually test for the existence of barriers to reproduction and apply the biological species concept. *Spialia sertorius* and *S. orbifer* are widely recognized parapatric species with a contact zone in central Europe (Lorković 1973; Fazekas 1986; Tshikolovets 2011). The manifest different background colour in their hindwing undersides (reddish vs. greenish) and the apparent absence of intermediate forms in the contact areas (Hesselbarth *et al.* 1995) support their status. Species delimitation methods supported the species status for *S. sertorius* and split *S. orbifer* into two to three entities (in addition to *S. rosae*). Although we do not have evidence for testing the hypothesis of further species within *S. orbifer* at present, this result indicates a direction for future research.

In the case of *S. rosae*, a key piece in the puzzle comes from its ecology: it apparently specializes on *Rosa* spp. as host plant, while the rest of taxa in the group specialize on *Sanguisorba* spp. We do not consider the use of different host plants alone as proof for species status, because many taxa widely recognized as infraspecific, and even populations, specialize ecologically to variable degrees (Mallet 2008). However, in sympatry this character can function in combination with others to test the existence of gene flow and speciation (Mallet 2008; Hernández-Vera *et al.* 2010; Nosil 2012). We can assume that host plant choice and mitochondrial markers are a priori independent traits and their 100% match at individual level in sympatry strongly suggests the existence of two species based on the biological species concept (see a similar case in McBride *et al.* 2009).

Infection by *Wolbachia* was another character that correlated perfectly with *COI* and host plant in sympatry: no infection was detected in any specimen of *S. sertorius*, while all the *S. rosae* specimens tested were infected. *Wolbachia* may constitute a partial barrier to gene flow because of cytoplasmic incompatibility, but such a barrier is usually not strong and lasting enough as to consider the populations involved as different species (e.g. Ritter *et al.* 2013). Nevertheless, it may still represent a barrier that promotes speciation through reinforcement, especially in combination with other factors. The fact that two different strains have been documented for *S. rosae*, including one specimen apparently infected by both simultaneously, further complicates the interpretation of the effect of *Wolbachia* on gene flow between *S. sertorius* and *S. rosae*. Moreover, *Wolbachia*, like the mitochondrial DNA, is maternally inherited, and thus, it is not surprising that a correlation between these two markers exists. Thus, in no case we rely on infection by *Wolbachia* in order to take taxonomic decisions, but in order to document potential factors at play in the speciation process of these taxa.

While most species delimitation analyses supported *S. rosae* as a species, we must acknowledge that the exact phylogenetic position of this taxon is not fully understood: depending on the markers, this taxon is very close to *S. sertorius* (*ITS2* and morphology), or more closely related to *S. orbifer* (1.1% minimum *COI* p-distance to *S. orbifer*) (Fig. 4). Discordance among markers is a common phenomenon that can be produced by lineage sorting effects or, alternatively, by introgression. According to the phylogenetic relationships based on *COI S. orbifer* is paraphyletic because of *S. rosae* (Fig. S2, Supporting information). How can the close mitochondrial DNA relationship between the Iberian endemic *S. rosae* and the east European *S. orbifer* be explained, given the geographical distance? Although the Italian Peninsula is completely occupied by *S. sertorius*, including the tip of Aspromonte, this species is replaced in Sicily by an isolated *S. orbifer* population. Similar distribution patterns were found in other species along the Italian Peninsula and neighbouring islands, and they have been suggested to be the result of relatively recent invasions of the Italian Peninsula from other European regions which replaced the ancestral populations on the mainland while, due to their isolation, ancestral insular populations have been preserved as relicts (Dapporto & Bruschini 2012; Dapporto *et al.* 2012). Thus, it can be hypothesized that the contact zone between the parapatric *S. sertorius* and *S. orbifer* was in the past closer to the Iberian Peninsula. There is evidence that hybridization and reinforcement in species' contact zones may result in new species (Mavárez *et al.* 2006; Mallet 2007) and it is possible that *S. rosae* was generated in that putative contact zone.

As most taxa in the subfamily Pyrginae seem to rely on chemical recognition (Hernández-Roldán *et al.* 2014), the composition of CHCs could be the proximate mechanism involved in the emergence of reproductive barriers also in the case of *S. rosae*. This hypothesis is supported by the fact that the cuticular hydrocarbon profile of *S. rosae* is the most differentiated in the group. It is well known that insect recognition systems do not only rely on genetically inherited cues but also on

ecologically acquired ones (e.g. Liang & Silverman 2000). Thus, the *S. rosae* chemical profile may either be a product of the host plant species used during development or generated independently by the butterfly, but in any case constitutes a potential way for females to avoid heterospecific mating. It has to be noted that the chemical signature of *S. rosae* is characterized by a particularly high frequency of alcohols, branched and unsaturated CHCs, as well as a relatively lower fraction of linear saturated hydrocarbons. According to a classic paradigm in insect chemical communication, carbon chains bearing double bounds and functional groups are more easily identified by receptors than linear alkanes (e.g. Breed 1998; Dani *et al.* 2005). From this perspective, *S. rosae* is characterized by a highly distinctive signature. If a prezygotic barrier mediated by chemical cues is established, wing pattern and genitalia shape could become neutral characters. Thus, the virtually identical morphology between *S. rosae* and *S. sertorius* may be the result of stasis of plesiomorphic characters, albeit other hypotheses such as homogenization through occasional introgressive hybridization cannot be ruled out.

Several studies showed that chromosomal rearrangements arising in primary contact zones could enhance the differentiation between populations and ultimately lead to ecological speciation (Feder *et al.* 2003a,b, 2005). A stable chromosome number ($n = 31$) found in all the studied species suggests that the genus *Spialia* is an example of chromosomal conservatism, where the plesiomorphic chromosome number ancestral for all heteroneuran Lepidoptera (Lukhtanov 2000) was preserved. This finding was unexpected because many other genera of Hesperiidae demonstrate chromosomal instability, a situation in which multiple closely related taxa (populations, subspecies and species) belonging to a single phylogenetic lineage drastically differ from each other by major chromosomal rearrangements resulting in a high variability of chromosome number (Lukhtanov 2014 and references therein). However, the uniformity of karyotypes does not imply that chromosome rearrangements were not involved in genome evolution. Numerous inter- or intrachromosomal rearrangements, such as translocations and inversions, can contribute to karyotype evolution without significant changes in chromosome number and size. The detection of these rearrangements is difficult in Lepidoptera due to the holocentric organization of their chromosomes (Carpenter *et al.* 2005; Vershinina *et al.* 2015). In our study, the FISH technique provided evidence for heterozygosity in the location of rDNA clusters, which is indicative of the occurrence of intrachromosomal rearrangements and/or gene movement in *S. sertorius*. In general, it seems that tandem arrays of major ribosomal RNA (rRNA) genes undergo dynamic evolution in Lepidoptera (Nguyen *et al.* 2010; Šíchová *et al.* 2013, 2015). However, several studies showed that the number and location of rDNA could be a useful marker for the study of karyotype evolution and the identification of a new species (Hirai *et al.* 1996; Roy *et al.* 2005; Bombarová *et al.* 2007). In *Spialia*, the rDNA analysis by itself did not confirm the differentiation of the species *S. sertorius* and *S. rosae*. However, it suggests ongoing chromosomal changes of unknown extent in the karyotype of *S. sertorius* that can be potentially important in ecological speciation and deserve further research. The *S. rosae–S. sertorius* system is an ideal model to study speciation linked to adaptation to novel trophic resources using emerging genomic techniques because they are recent sympatric species with apparently few other differences that could hamper the interpretation of the results.

Conclusion

Cryptic species represent a new dimension in the exploration of biodiversity (Bickford *et al.* 2007). Thus, while poorly studied areas and taxonomic groups still provide a wealth of morphologically well-differentiated species, the question arises on how many unnoticed species each form may hide. We must acknowledge that at this point it is unknown what fraction of total diversity is represented by cryptic species, both in general and for butterflies in particular. A recent estimate shows that 27% of the European butterflies include divergent mitochondrial lineages that highlight a considerable potential for cryptic taxa (Dincă *et al.* 2015). The results here presented exemplify the importance of comparing a variety of data sets through the so-called 'integrative approach'. The case of *S. rosae*, initially highlighted as a diverged mitochondrial lineage (Dincă *et al.* 2015), is paradigmatic because this species cannot be unambiguously distinguished from the sympatric *S. sertorius* based on the nuclear markers studied, adult morphometric analyses of external and internal organs, morphology of immature stages and karyotype, which would have led most researchers to conclude that a single species exists. Nevertheless, novel host plant data, CHC composition and infection by *Wolbachia* differentiated the two taxa in addition to *COI*. Importantly, none of these characters independently constitute in principle definitive proof of specific status. Only the exact match at specimen level of these three characters (primarily of *COI* and host plant, as the chemical character could be a direct outcome of host plant) in sympatry prove that two species exist. This study represents yet another example that the integrative approach keeps on yielding new species even in intensively studied regions, and

contributes to an increased awareness that global species richness is probably underestimated.

Description of *Spialia rosae* sp. nov.

Spialia rosae Hernández-Roldán, Dapporto, Dincă, Vicente & Vila sp. nov.

Type material (Appendix S2 Fig. A1. 1–16, Supporting information).

Holotype: Male (Appendix S2 Fig. A1. 1, Supporting information). Puerto de la Ragua, Sierra Nevada, Granada, Spain, 2090 m, 37.1108, −3.0344, 17.vii.2011 (J. Hernández-Roldán, V. Dincă, J. C. Vicente & R. Vodă leg.; Ex coll. J. Hernández-Roldán 6608; In coll. Museo Nacional de Ciencias Naturales, Madrid, Spain; Tissues in coll. Institut de Biologia Evolutiva (CSIC-UPF), Barcelona, Spain, under code 11-G137; Karyotype preparation in the Department of Karyosystematics, Zoological Institute of Russian Academy of Sciences, St. Petersburg, Russia. *COI* GenBank Accession no.: KU905538, *ITS2* GenBank Accession no.: KU905636, *Wg* GenBank Accession no.: KU905694.

Paratypes: Nine eggs, 31 larvae, 57 males and 10 females, all from Spain (Appendix S2 Fig. A1. 2–16, Supporting information). See collecting and repository data, and GenBank accession numbers in Table S1 and in Appendix S2 Table A1 (Supporting information).

Diagnosis: Spialia rosae sp. nov. can be distinguished from the most closely related species (*S. sertorius*, *S. orbifer*, *S. ali* and *S. therapne*) on the basis of the DNA sequence of the mitochondrial gene cytochrome c oxidase subunit I (*COI)*, the composition of cuticular hydrocarbons on the wings, and the host plant preference. A univocal mitochondrial diagnostic character is represented by a cytosine (C) in position 193 in *S. rosae* sp. nov. *COI* mtDNA (positions refer to the Holotype sequence, Genbank Accession no. KU905538). *Spialia rosae* sp. nov. is differentiated from the sympatric and synmorphic *S. sertorius* based on the following characters in *COI*: thymine (T) or cytosine (C) in position 43 vs. adenine (A) or guanine (G) in *S. sertorius;* T vs. C in position 49; A or G vs. T in position 82; C vs. T in 103; T vs. C in 124; C vs. T in 187; C vs. T in 193; C vs. T in 226; T vs. C in 271; C vs. T in 283; T vs. A in 298; C vs. T in 340; A vs. C in 364; T vs. C in 400; and T vs. A in 412, respectively. Cuticular hydrocarbons on the wings are characterized by the abundance of C27:1, C29:1, 7methylC27, central-methylC26 and 11,Y dimethylC25, unlike the rest of studied species. Moreover, in *S. rosae* sp. nov. shows the lowest abundances (near to zero) of C25:1, 2methylC26 and 2methylC28, compounds that are well represented in *S. sertorius* and *S. orbifer* (Fig. 4). *Spialia rosae* feeds in nature on *Rosa* spp. instead of *Sanguisorba* spp. in other species. Current data indicate that no morphological character allows for a reliable separation between *S. rosae* sp. nov. and *S. sertorius.* See details of morphology, DNA markers, karyotype, cuticular hydrocarbons on the wings, *Wolbachia* infection, host plants, etymology, distribution and remarks in Appendix S2 (Supporting information).

Acknowledgements

We are grateful to the colleagues who provided samples for this study: J. Estela; E. García-Barros, S. Kunze, M. L. Munguira, J. G. Renom, J. Requejo, S. Scalercio, S. Viader and R. Vodă. B. Emerson, N. Wahlberg and six anonymous reviewers provided helpful discussions and comments. Special thanks are given to L. Sáez and J. Tapia for determining the plants, to E. Brockmann for help in butterfly determination, and to G. Lamas for advice on the species status and description. Funding for this research was provided by the Spanish Ministerio de Economía y Competitividad (Project CGL2013-48277-P), by the projects 'Barcoding Italian Butterflies', by two Marie Curie International Outgoing Fellowships within the 7th European Community Framework Programme to V. Dincă (project no. 625997) and to E. Hornett (project no. 330136) and by European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Grant (project no. 658844) to L. Dapporto. J. Šíchová was supported by Grant 14-22765S of the Czech Science Foundation, and V. Lukhtanov by RFBR grants 15-29-02533, 15-04-01581, 14-04-00139, 14-04-01051 and 14-04-00770.

References

Aitchison J (1986) *The Statistical Analysis of Compositional Data*. Chapman & Hall, London.

Andújar C, Serrano J, Gómez-Zurita J (2012) Winding up the molecular clock in the genus *Carabus* (Coleoptera: Carabidae): assessment of methodological decisions on rate and node age estimation. *BMC Evolutionary Biology*, **12**, 40.

Baldo L, Dunning Hotopp JC, Jolley KA *et al.* (2006) Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Applied and Environmental Microbiology*, **72**, 7098–7110.

Balletto E, Cassulo LA, Bonelli S (2014) An annotated checklist of the Italian butterflies and skippers (Papilionoidea, Hesperiioidea). *Zootaxa*, **3853**, 1–114.

Beheregaray LB, Caccone A (2007) Cryptic biodiversity in a changing world. *Journal of Biology*, **6**, 9.

Bickford D, Lohman DJ, Sodhi NS *et al.* (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, **22**, 148–155.

Bombarová M, Marec F, Nguyen P, Špakulová M (2007) Divergent location of ribosomal genes in chromosomes of fish thorny-headed worms, *Pomphorhynchus laevis* and *Pomphorhynchus tereticollis* (Acanthocephala). *Genetica*, **131**, 141–149.

Bookstein FL (1991) *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, Cambridge, UK, New York.

Breed MD (1998) Recognition pheromones of the honey bee. *BioScience*, **48**, 463–470.

Britton T, Anderson C, Jacquet D, Lundqvist S, Bremer K (2006) PATHd8 – a program for phylogenetic dating of large

trees without a molecular clock. Available at www.-math.su.se/PATHd8/

Brower AVZ (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial-DNA evolution. *Proceedings of the National Academy of Sciences USA*, **91**, 6491–6495.

Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PDN (2008) DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservación Guanacaste, Costa Rica. *Proceedings of the National Academy of Sciences USA*, **105**, 6350–6355.

Carpenter JE, Bloem S, Marec F (2005) Inherited sterility in insects. In: *Sterile Insect Technique. Principles and Practice in Area-Wide Integrated Pest Management*(eds Dyck VA, Hendrichs J, Robinson AS), pp. 115–146. Springer, Dordrecht.

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

Dani FR, Jones GR, Corsi S, Beard R, Pradella D, Turillazzi S (2005) Nestmate recognition cues in the honey bee: differential importance of cuticular alkanes and alkenes. *Chemical Senses*, **30**, 477–489.

Dapporto L, Bruschini C (2012) Invading a refugium: post glacial replacement of the ancestral lineage of a Nymphalid butterfly in the West Mediterranean. *Organisms Diversity & Evolution*, **12**, 39–49.

Dapporto L, Bruschini C, Dincă V, Vila R, Dennis RLH (2012) Identifying zones of phenetic compression in West Mediterranean butterflies (Satyrinae): refugia, invasion and hybridization. *Diversity and Distributions*, **18**, 1066–1076.

Dapporto L, Vodă R, Dincă V, Vila R (2014a) Comparing population patterns for genetic and morphological markers with uneven sample sizes. An example for the butterfly *Maniola jurtina*. *Methods in Ecology and Evolution*, **5**, 834–843.

Dapporto L, Fattorini S, Vodă R, Dincă V, Vila R (2014b) Biogeography of western Mediterranean butterflies: combining turnover and nestedness components of faunal dissimilarity. *Journal of Biogeography*, **41**, 1639–1650.

De Jong R (1974) Systematics and evolution of the Palaearctic *Spialia* species (Lepidoptera, Hesperiidae). *Tijdschrift Voor Entomologie*, **117**, 225–271.

De Jong R (1978) Monograph of the genus *Spialia* Swinhoe (Lepidoptera, Hesperiidae). *Tijdschrift Voor Entomologie*, **121**, 23–146.

Dincă V, Dapporto L, Vila R (2011a) A combined genetic-morphometric analysis unravels the complex biogeographical history of *Polyommatus icarus* and *Polyommatus celina* Common Blue butterflies. *Molecular Ecology*, **20**, 3921–3935.

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

Dincă V, Zakharov EV, Hebert PDN, Vila R (2011c) Complete DNA barcode reference library for a country's butterfly fauna reveals high performance for temperate Europe. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 347–355.

Dincă V, Wiklund C, Lukhtanov VA *et al.* (2013) Reproductive isolation and patterns of genetic differentiation in a cryptic butterfly species complex. *Journal of Evolutionary Biology*, **26**, 2095–2106.

Dincă V, Montagud S, Talavera G *et al.* (2015) DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. *Scientific Reports*, **5**, 12395.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Excoffier L, Langaney A (1989) Origin and differentiation of human mitochondrial DNA. *American Journal of Human Genetics*, **44**, 73–85.

Fazekas I (1986) Die *Spialia*-Arten des Karpatenbeckens und ihre Verbreitung (Lepidoptera: Hesperiidae). *Nachrichten des Entomologischen Vereins Apollo*, **7**, 49–55.

Feder JL, Berlocher SH, Roethele JB *et al.* (2003a) Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proceedings of the National Academy of Sciences USA*, **100**, 10314–10319.

Feder JL, Roethele JB, Filchak K, Niedbalski J, Romero-Severson J (2003b) Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. *Genetics*, **163**, 939–953.

Feder JL, Xie XF, Rull J *et al.* (2005) Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in *Rhagoletis*. *Proceedings of the National Academy of Sciences USA*, **102**, 6573–6580.

Feulner PGD, Kirschbaum F, Schugardt C, Ketmaier V, Tiedemann R (2006) Electrophysiological and molecular genetic evidence for sympatrically occurring cryptic species in African weakly electric fishes (Teleostei: Mormyridae: Campylomormyrus). *Molecular Phylogenetics and Evolution*, **39**, 198–208.

Fujisawa T, Barraclough TG (2013) Delimiting species using single-locus data and the generalized mixed yule coalescent approach: a revised method and evaluation on simulated data sets. *Systematic Biology*, **65**, 707–724.

Fuková I, Nguyen P, Marec F (2005) Codling moth cytogenetics: karyotype, chromosomal location of rDNA and molecular differentiation of sex chromosomes. *Genome*, **48**, 1083–1092.

Giraldo N, Salazar C, Jiggins CD, Bermingham E, Lianres M (2008) Two sisters in the same dress: *Heliconius* cryptic species. *BMC Evolutionary Biology*, **8**, 324.

Grundt HH, Kjolner S, Borgen L, Rieseberg LH, Brochmann C (2006) High biological species diversity in the arctic flora. *Proceedings of the National Academy of Sciences USA*, **103**, 972–975.

Hausmann A, Haszprunar G, Segerer AH, Speidel W, Behounek G, Hebert PDN (2011) Now DNA-barcoded: the butterflies and larger moths of Germany. *Spixiana*, **34**, 47–58.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences USA*, **101**, 14812–14817.

Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, **487**, 94–98.

Hernández-Roldán JL, Bofill R, Dapporto L, Munguira ML, Vila R (2014) Morphological and chemical analysis of male scent organs in the butterfly genus *Pyrgus* (Lepidoptera: Hesperiidae). *Organisms Diversity and Evolution*, **14**, 269–278.

Hernández-Vera G, Mitrović M, Jović J *et al.* (2010) Host-associated genetic differentiation in a seed parasitic weevil *Rhinusa antirrhini* (Coleptera: Curculionidae) revealed by mitochondrial and nuclear sequence data. *Molecular Ecology*, **19**, 2286–2300.

Hesselbarth G, Van Oorschot H, Wagener S (1995) *Die Tagfalter der Türkei unter Berücksichtigung der angrenzenden Länder*, Vol. 1–3. Selbstverlag S. Wagener, Bocholt.

Hirai H, Yamamoto MT, Taylor RW, Imai HT (1996) Genomic dispersion of 28S rDNA during karyotypic evolution in the ant genus Myrmecia (Formicidae). *Chromosoma*, **105**, 190–196.

Kodandaramaiah U, Wahlberg N (2009) Phylogeny and biogeography of Coenonympha butterflies (Nymphalidae: Satyrinae)–patterns of colonization in the Holarctic. *Systematic Entomology*, **34**, 315–323.

Kudrna O, Pennerstorfer J, Lux K (2015) Distribution atlas of European Butterflies and Skippers. Wissenschaftlicher.

Lanfear R, Calcott B, Ho SY, Guindon S (2012) Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.

Lê Cao KA, Boitard S, Besse P (2011) Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics*, **12**, 253.

Leigh JW, Susko E, Baumgartner M, Roger AJ (2008) Testing congruence in phylogenomic analysis. *Systematic Biology*, **57**, 104–115.

Liang D, Silverman J (2000) "You are what you eat": diet modifies cuticular hydrocarbons and nestmate recognition in the Argentine ant, *Linepithema humile*. *Naturwissenschaften*, **87**, 412–416.

Lorković Z (1973) 150 Jahre bis zur Entdeckung der Präimaginalstadien von *Spialia orbifer* Hbn. (Lep., Hesperiidae). *Acta Entomologica Yugoslavica*, **9**, 67–70.

Lukhtanov VA (2000) Sex chromatin and sex chromosome systems in nonditrysian Lepidoptera (Insecta). *Journal of Zoological Systematics and Evolutionary Research*, **38**, 73–79.

Lukhtanov VA (2014) Chromosome number evolution in skippers (Lepidoptera, Hesperiidae). *Comparative Cytogenetics*, **8**, 275–291.

Lukhtanov VA, Shapoval NA, Dantchenko AV (2008) *Agrodiaetus shahkuhensis* sp. n. (Lepidoptera, Lycaenidae), a cryptic species from Iran discovered by using molecular and chromosomal markers. *Comparative Cytogenetics*, **2**, 99–114.

Lukhtanov VA, Dantchenko AV, Vishnevskaya MS, Saifitdinova AF (2015) Detecting cryptic species in sympatry and allopatry: analysis of hidden diversity in *Polyommatus* (*Agrodiaetus*) butterflies (Lepidoptera: Lycaenidae). *Biological Journal of the Linnean Society*, **116**, 468–485.

Mallet J (2007) Hybrid speciation. *Nature*, **446**, 279–283.

Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **363**, 2971–2986.

Mavárez J, Salazar CA, Bermingham E, Salcedo C, Jiggins CD, Linares M (2006) Speciation by hybridization in Heliconius butterflies. *Nature*, **441**, 868–871.

McBride CS, Velzen RV, Larsen TB (2009) Allopatric origin of cryptic butterfly species that were discovered feeding on distinct host plants in sympatry. *Molecular Ecology*, **18**, 3639–3651.

Mediouni J, Fuková I, Frydrychová R, Dhouibi MH, Marec F (2004) Karyotype, sex chromatin and sex chromosome differentiation in the carob moth, *Ectomyelois ceratoniae* (Lepidoptera: Pyralidae). *Caryologia*, **57**, 184–194.

Nguyen P, Sahara K, Yoshido A, Marec F (2010) Evolutionary dynamics of rDNA clusters on chromosomes of moths and butterflies (Lepidoptera). *Genetica*, **138**, 343–354.

Nosil P (2012) *Ecological Speciation*. Oxford University Press, Oxford.

Pons J, Barraclough TG, Gomez-Zurita J *et al.* (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595–610.

Posada D (2008) JModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

Puillandre N, Cruaud C, Kantor YI (2010) Cryptic species in *Gemmuloborsonia* (Gastropoda: Conoidea). *Journal of Molluscan Studies*, **76**, 11–23.

Quek SP, Davies SJ, Ashton PS, Itino T, Pierce NE (2004) Codiversification in an ant-plant mutualism: stem texture and the evolution of host use in *Crematogaster* (Formicidae: Myrmicinae) inhabitants of *Macaranga* (Euphorbiaceae). *Evolution*, **58**, 554–570.

Ritter S, Michalski SG, Settele J *et al.* (2013) *Wolbachia* infections mimic cryptic speciation in two parasitic butterfly species, *Phengaris teleius* and *P. nausithous* (Lepidoptera: Lycaenidae). *PLoS One*, **8**, e78107.

Ronquist F, Teslenko M, van der Mark P *et al.* (2012) MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Rowe CL, Adamowicz SJ, Hebert PDN (2007) Three new cryptic species of the freshwater zooplankton genus *Holopedium* (Crustacea: Branchiopoda: Ctenopoda), revealed by genetic methods. *Zootaxa*, **1656**, 1–49.

Roy V, Monti-Dedieu L, Chaminade N *et al.* (2005) Evolution of the chromosomal location of rDNA genes in two *Drosophila* species subgroups: *ananassae* and *melanogaster*. *Heredity*, **94**, 388–395.

Sanders KL, Malhotra A, Thorpe RS (2006) Combining molecular, morphological and ecological data to infer species boundaries in a cryptic tropical pitviper. *Biological Journal of the Linnean Society*, **87**, 343–364.

Schmitt T (2007) Molecular biogeography of Europe: pleistocene cycles and postglacial trends. *Frontiers in Zoology*, **4**, 11.

Shiraiwa K, Cong Q, Grishin NV (2014) A new *Heraclides* swallowtail (Lepidoptera, Papilionidae) from North America is recognized by the pattern on its neck. *ZooKeys*, **468**, 85–135.

Šíchová J, Nguyen P, Dalíková M, Marec F (2013) Chromosomal evolution in tortricid moths: conserved karyotypes with diverged features. *PLoS One*, **8**, e64520.

Šíchová J, Voleníková A, Dincă V *et al.* (2015) Dynamic karyotype evolution and unique sex determination systems in *Leptidea* wood white butterflies. *BMC Evolutionary Biology*, **15**, 89.

Smith AM, Rodriguez JJ, Whitfield JB *et al.* (2008) Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. *Proceedings of the National Academy of Sciences USA*, **105**, 12359–12364.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

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

Tolman T, Lewington R (2008) *Collins Butterfly Guide*. Harper-Collins, London.

Tshikolovets VV (2011) *Butterflies of Europe & the Mediterranean Area*. Tshikolovets Publications, Pardubice.

Vershinina AO, Lukhtanov VA (2010) Geographical distribution of the cryptic species *Agrodiaetus alcestis alcestis*, *A. alcestis karacetinae* and *A. demavendi* (Lepidoptera, Lycaenidae) revealed by cytogenetic analysis. *Comparative Cytogenetics*, **4**, 1–11.

Vershinina AO, Anokhin BA, Lukhtanov VA (2015) Ribosomal DNA clusters and telomeric (TTAGG)n repeats in blue butterflies (Lepidoptera, Lycaenidae) with low and high chromosome numbers. *Comparative Cytogenetics*, **9**, 161–171.

Vodă R, Dapporto L, Dincă V, Vila R (2015) Why do cryptic species tend not to co-occur? A case study on two cryptic pairs of butterflies. *PLoS One*, **10**, e0117802.

Vrijenhoek RC, Schutz SJ, Gustafson RG, Lutz RA (1994) Cryptic species of deep sea clams (Mollusca, Bivalvia, Vesicomyidae) from hydrothermal vent and cold water seep environments. *Deep Sea Research Part I*, **41**, 1171–1189.

Zemlak TS, Ward RD, Connell AD, Holmes BH, Hebert PDN (2009) DNA barcoding reveals overlooked marine fishes. *Molecular Ecology Resources*, **9**, 237–243.

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

Zinetti F, Dapporto L, Vovlas A *et al.* (2013) When the rule becomes the exception. No evidence of gene flow between two *Zerynthia* cryptic butterflies suggests the emergence of a new model group. *PLoS One*, **8**, e65746.

J.L.H.R., L.D., R.V. and V.D. designed the study and wrote a first draft of the manuscript; J.C.V., J.L.H.R., L.D., R.V. and V.D. collected specimens and ecological data on the field; J.S. and V.L. carried out cytogenetic analyses; E.A.H. carried out *Wolbachia* analyses; V.D. and G.T. carried out genetic analyses (DNA extraction, sequencing and phylogenetic trees); L.D., R.V. and V.D. performed chemical analyses; J.L.H.R., L.D. and V.D. carried out geometric morphometrics and performed multivariate and GIS analyses; all authors contributed to the writing of the MS.

Data accessibility

DNA sequences: GenBank accession numbers are listed in Table S1 (Supporting information); Barcode of Life Data Systems data set DS-SPIALIA (DOI: dx.doi.org/10.5883/DS-SPIALIA). Individual sample data: Table S1 (Supporting information).

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 Additional methods.

Appendix S2 Additional data for the description of *Spialia rosae* sp. nov.

Table S1 List of samples used.

Table S2 Chromosome number results.

Table S3 Frequency of *Wolbachia* infection in the studied specimens.

Table S4 *Spialia sertorius* and *Spialia rosae* sp. nov. host plant results.

Fig. S1 Location of fixed landmarks (green dots) and sliding semilandmarks (white dots) on hindwings (left) and valvae (right).

Fig. S2 Bayesian phylogeny based on *COI* haplotypes.

Fig. S3 Entities of the *Spialia sertorius* species group recovered by the single-threshold Generalized Mixed Yule-Coalescent model (GMYC), and Poisson Tree Processes (PTP and bPTP), based on 1009 *COI* Pyrginae sequences.

Fig. S4 Geographical distribution of mitochondrial *COI* haplotypes for *S. ali* stat. rest. (red) and *S. therapne* stat. rest. (green).

Fig. S5 Bayesian phylogeny based on *ITS2* sequences.

Fig. S6 Bayesian phylogeny based on *Wg* sequences.

Fig. S7 Bayesian phylogeny based on *COI* and *Wg* sequences.

Fig. S8 Entities of the *Spialia sertorius* species group recovered by the single-threshold Generalized Mixed Yule-Coalescent model (GMYC), and Poisson Tree Processes (PTP and bPTP), based on the combined data set of *COI*, *ITS2* and *Wg* DNA sequences.

Fig. S9 Bayesian chronogram based on *COI*, *Wg* and *ITS2* sequences.

Fig. S10 Histogram showing the average transformed relative abundance of the 23 detected cuticular compounds for each species.

Fig. S11 Karyotype results.

Fig. S12 Localization of rDNA clusters in spread chromosome preparations of three *Spialia* species by fluorescence in situ hybridization (FISH) with 18S rDNA probe.

Fig. S13 Scatterplot of the first two PLSDA components for wing shape and pattern.

Fig. S14 Scatterplot of the two most important relative warps in determining the diversification among taxa (PC1 and PC26) in hindwings.

Fig. S15 Scatterplot of the first two PLSDA components for genitalia.

Fig. S16 Scatterplot of the two most important relative warps in determining the diversification among taxa (PC1 and PC6).

Fig. S17 Tree of *Wolbachia* strains.