Zoological Journal of the Linnean Society, 2020, **XX**, 1–13. With 4 figures.

Overlooked cryptic diversity in *Muschampia* (Lepidoptera: Hesperiidae) adds two species to the European butterfly fauna

JOAN C. HINOJOSA[1], LEONARDO DAPPORTO[2], ERNST BROCKMANN[3], VLAD DINCĂ[4], VALENTIN TIKHONOV[5], NICK GRISHIN[6,7,8], VLADIMIR A. LUKHTANOV[9] and ROGER VILA[1,*]

[1]*Institut de Biologia Evolutive (CSIC-UPF), Passeig de la Barceloneta 37-49, 08003, Barcelona, Spain*
[2]*ZEN lab, Dipartimento di Biologia, University of Florence, via Madonna del Piano 6, 50019 Sesto Fiorentino, Italy*
[3]*Laubacher Straße 4, 35423 Lich, Hessen, Germany*
[4]*Ecology and Genetics Research Unit, PO Box 3000, University of Oulu, 90014, Finland*
[5]*North Caucasus Federal University, Ulitsa Pushkina 1, Stavropol 355009, Russia*
[6]*Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, MD 20815, USA*
[7]*Department of Biophysics, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-8816, USA*
[8]*Department of Biochemistry, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-9038, USA*
[9]*Department of Karyosystematics, Zoological Institute of the Russian Academy of Sciences, Universitetskaya Naberezhnaya 1, St. Petersburg 199034, Russia*

Received 14 May 2020; revised 2 November 2020; accepted for publication 9 November 2020

Cryptic species represent a challenge for documenting global biodiversity. Even in well-studied groups, such as European butterflies, the application of integrative approaches has allowed the recognition of an unexpected number of cryptic taxa. Here, we combine the analysis of mitochondrial (cytochrome *c* oxidase I, *COI*) and nuclear (internal transcribed spacer 2, ITS2) markers with geometric morphometrics of the male genitalia to study diversity within the butterfly *Muschampia proto*. The nuclear marker reveals three well-supported and deeply diverged lineages, which are also detected based on mitochondrial DNA, although the latter recovers one of them as paraphyletic with poor support. These lineages also present distinct male genital characters, which allow blind assignment of > 97% of specimens when applying a jackknife procedure. We conclude that *M. proto* comprises three cryptic species that started to differentiate ~2 Mya: *M. proto*, distributed in northern Africa, the Iberian Peninsula and southern France; ***Muschampia alta* comb. & stat. nov.**, occurring in southern Italy and the Balkan Peninsula; and *Muschampia proteides*, present in the easternmost part of Europe, the Near East and Iran. This discovery adds two new species to the European butterfly fauna and highlights the necessity to continue investigating potential cryptic diversity.

ADDITIONAL KEYWORDS: cryptic species – morphometrics – phylogenetics – speciation – taxonomy.

INTRODUCTION

The increasing number of cryptic species discovered in recent years (Struck *et al.*, 2018) demonstrates that our knowledge of biodiversity is still incomplete. This limitation can lead to key taxa in need of conservation being overlooked and can hamper a comprehensive understanding of evolutionary processes and ecosystem functioning (Bickford *et al.*, 2007; Alizon *et al.*, 2008; Ceballos & Ehrlich, 2009; Esteban & Finlay, 2010). Other than representing a substantial fraction of alpha diversity (Bickford *et al.*, 2007; Pfenninger & Schwenk, 2007), cryptic species have been shown to encompass a considerable percentage of regional beta

*Corresponding author. E-mail: roger.vila@ibe.upf-csic.es

© 2021 The Linnean Society of London,

Downloaded from https://academic.oup.com/zoolinnean/advance-article/doi/10.1093/zoolinnean/zlaa171/6069183 by guest on 08 January 2021

diversity, owing to their tendency to show allopatric or parapatric distributions (Vodă *et al.*, 2015). In relatively well-studied taxa, such as European butterflies, nearly 28% of the currently accepted species display deeply diverged mitochondrial intraspecific lineages (Dincă *et al.*, 2015), some of which are concentrated in particular geographical regions (Dapporto *et al.*, 2017; Scalercio *et al.*, 2020). These taxa should be analysed in detail to ascertain the existence of potential cryptic species. In fact, new cryptic butterfly species have recently been documented in Europe (Dincă *et al.*, 2011a, b; Zinetti *et al.*, 2013; Hernández-Roldán *et al.*, 2016; Schmitt *et al.*, 2016; Verovnik & Wiemers, 2016; Lukhtanov & Dantchenko, 2017). Nevertheless, numerous taxa showing highly diverging lineages remain largely unexplored, hampering a thorough evaluation of the incidence of cryptic biodiversity in Lepidoptera.

The genus *Muschampia* Tutt, 1906 or *Sloperia* Tutt, 1906 (the validity of the denomination is still under debate; Wiemers *et al.*, 2018) comprised, in its former taxonomy, ~20 species distributed across the Palaearctic region, with most diversity occurring in the Middle East and Central Asia (Tuzov, 1997; García-Barros *et al.*, 2013). Recently, the genus has been revised and rearranged based on genomic analyses (Zhang *et al.*, 2020). It has been shown that *Muschampia* is not monophyletic, as traditionally accepted. To re-establish monophyly, *Muschampia cribrellum* (Eversmann, 1841) has been excluded from the genus (now *Favria cribrellum*), and several species previously considered as members of the genus *Carcharodus* Hübner [1819] have been included in the genus *Muschampia* (Zhang *et al.*, 2020).

Muschampia proto (Ochsenheimer, 1808) is the type species of the genus. It was believed to range from the Mediterranean (northern Africa, Iberian Peninsula, southern France, southern Italy, southern Balkan Peninsula, Anatolia, Lebanon and Israel) to areas surrounding the Caucasus. Its distribution is fragmented across the range, which led to the description of several subspecific taxa (Tshikolovets, 2011). The eastern taxon *Muschampia proteides* (Wagner, 1929), originally described as a subspecies of *M. proto*, is frequently treated as a different species (e.g. Tshikolovets, 2011; Kemal & Koçak, 2017) but, depending on the author, a set of subspecies are placed either in *M. proto* or in *M. proteides* (e.g. *M. p. sovietica* Sichel, 1964; *M. p. hieromax* Hemming, 1932). These taxonomic decisions have never been verified by pairing genetic data with detailed morphological analyses, such as geometric morphometrics. Given these uncertainties, the taxa related to *M. proto* and *M. proteides* were grouped under the denomination '*M. proto* complex'.

The so-called integrative approach (the analysis of presumably independent characters of different types) can provide better support for species delimitation compared with traditional morphological approaches or single-marker surveys (Lumley & Sperling, 2010; Dincă *et al.*, 2011a, b; Platania *et al.*, 2020). Accordingly, the main criterion used for species recognition in the most recent list of European butterflies (Wiemers *et al.*, 2018) was differentiation in two character sets (e.g. mitochondrial DNA, nuclear DNA, morphology or karyology). In this context, we studied the genetic structure of the *M. proto* complex using mitochondrial (cytochrome *c* oxidase I, *COI*) and nuclear (internal transcribed spacer 2, ITS2) markers, and we examined the morphology of male genitalia by applying geometric morphometrics. Our aims were to provide a better understanding of phylogeographical patterns and to refine the taxonomic status of this complex.

MATERIAL AND METHODS

Sampling, DNA extraction and sequencing

A total of 147 samples were used for the DNA analysis and/or measurement of the genitalia (Supporting Information, Table S1; Fig. 1). The phylogenetic analyses were conducted for 122 individuals, from which we retrieved 121 *COI* and 53 ITS2 sequences. A subset of the *COI* sequences was generated at the Biodiversity Institute of Ontario (Guelph, ON, Canada) following standard protocols for DNA barcoding (deWaard *et al.*, 2008), and DNA sequencing was performed on an ABI 3730XL capillary sequencer (Applied Biosystems). The rest of the *COI* sequences and all ITS2 sequences generated for this study were obtained at the Butterfly Diversity and Evolution Lab of the Institut de Biologia Evolutiva (CSIC-UPF; Barcelona, Spain). In this case, total genomic DNA was extracted using Chelex 100 resin, 100–200 mesh, sodium form (BioRad), using the following protocol: one leg was removed and introduced into 100 µL of 10% Chelex to which 5 µL of Proteinase K (20 mg/mL) was added. The samples were incubated overnight at 55 °C in a shaker and subsequently incubated at 100 °C for 15 min.

LepF1 and LepR1 primers (Hebert *et al.*, 2004; 5′-ATT CAACCAATCATAAAGATATTGG-3′ and 5′-TAAACT TCTGGATGTCCAAAAAATCA-3′, respectively) were used for the amplification of the standard barcode fragment of *COI*. Double-stranded DNA was amplified in 25 µL reactions: 13.2 µL ultra-pure (HPLC quality) water, 5 µL of 5× Green GoTaq Flexi Buffer (Promega), 3.2 µL of 25 mM $MgCl_2$, 0.5 µL of 10 mM dNTP, 0.5 µL of each primer (10 mM), 0.1 µL of GoTaq G2 Flexi Polymerase (Promega) and 2 µL of extracted DNA. The reaction conditions comprised a first denaturation at 92 °C for 60 s, followed by five cycles of 92 °C for 15 s, 48 °C for 45 s and 62 °C for 150 s, another 30 cycles with the annealing temperature changed to 52 °C, and a final extension step at 62 °C for 7 min.

Downloaded from https://academic.oup.com/zoolinnean/advance-article/doi/10.1093/zoolinnean/zlaa171/6069183 by guest on 08 January 2021

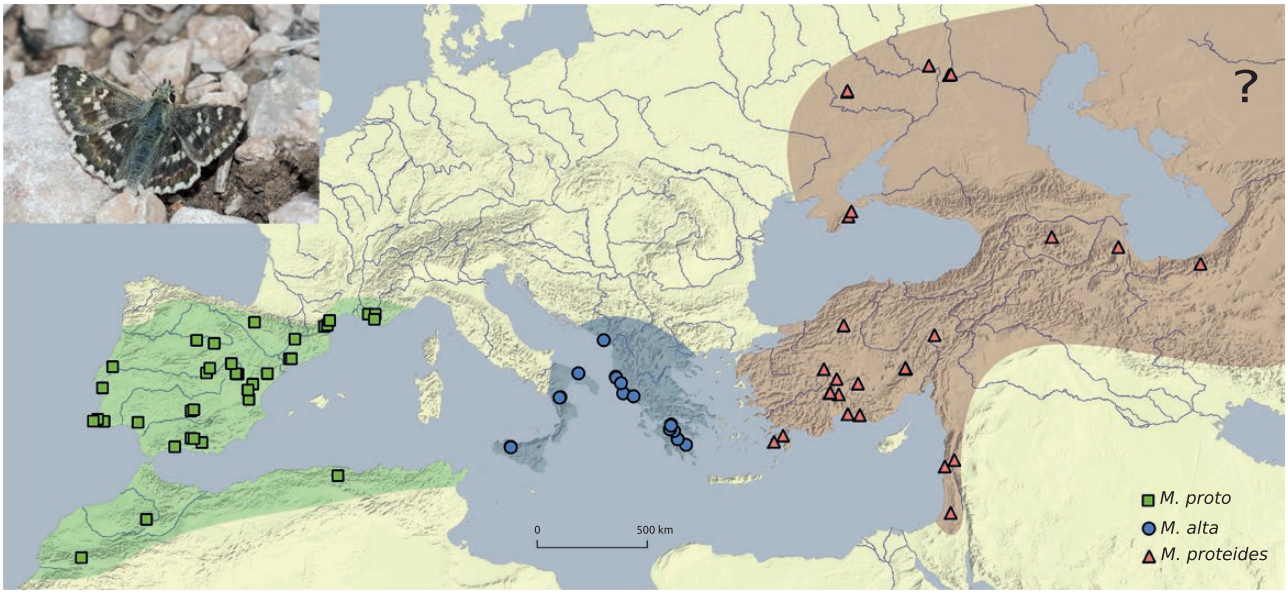

Figure 1. Sampling sites (symbols) and the approximate range (shading) of the species recognized in this study (green squares, *Muschampia proto*; blue circles, *Muschampia alta*; red triangles *Muschampia proteides*). Photograph: *M. proto* from Jaén (southern Iberia) by V.D.

ITS2 was amplified with MS-2F and MS-2R primers (5′-GGACACATTTGAACATCGACA-3′ and 5′-TGATCTGAGGCCAACGATAAA-3′, respectively), which we developed specifically for *Muschampia*. Double-stranded DNA was amplified in 25 µL reactions: 14.4 µL ultra-pure (HPLC quality) water, 5 µL of 5× Green GoTaq Flexi Buffer (Promega), 2 µL of 25 mM $MgCl_2$, 0.5 µL of 10 mM dNTP, 0.5 µL of each primer (10 mM), 0.1 µL of GoTaq G2 Flexi Polymerase (Promega) and 2 µL of extracted DNA. The reaction conditions comprised a first denaturation at 94 °C for 3 min, 36 cycles of 94 °C for 35 s, 48 °C for 30 s and 72 °C for 60 s, and a final extension step at 72 °C for 6 min. Polymerase chain reaction products were purified and Sanger sequenced by Macrogen Inc. Europe (Amsterdam, The Netherlands). All sequences are deposited in BOLD and GenBank (Supporting Information, Table S1). The mitochondrial *COI* sequence of the *Muschampia mohammed caid* (Le Cerf, 1923) holotype specimen was obtained from a previously published study (Zhang *et al.*, 2020).

Phylogenetic reconstruction

Sequences belonging to each marker were aligned with GENEIOUS v.11.0.5 (Kearse *et al.*, 2012) with the Geneious Alignment method. According to jModelTest v.2.1.7 (Darriba *et al.*, 2012) and under the Akaike information criterion, the best-fitting substitution model was GTR+G. A Bayesian inference phylogeny was constructed for each marker (Fig. 2A, B) with BEAST v.2.5.0 (Bouckaert *et al.*, 2014). The *COI* phylogeny was also used to obtain rough estimates of node ages, taking into account two molecular clocks: 1.5% uncorrected pairwise distance per million years, estimated for various invertebrates (Quek *et al.*, 2004), and 2.3%, estimated for the entire mitochondrial genome of several arthropods (Brower, 1994). A strict clock and a normal prior distribution were used, the latter centred on the mean between the two substitution rates, and the standard deviation was tuned so that the 95% confidence interval of the posterior density coincided with the 1.5 and 2.3% rates. For both markers, parameters were estimated using two independent runs of 30 million generations each (then merged), and convergence was checked with TRACER v.1.7.1 (Rambaut, 2018). Two maximum likelihood (ML) inferences were obtained, one for each gene, with RaxML v.8.2.12 (Stamatakis, 2014) in CIPRES (Miller *et al.*, 2015). The ML + thorough bootstrap workflow was selected, with a GTRGAMMA model and 1000 bootstrap replicates. Uncorrected *p*-distances (minimum interspecific and maximum intraspecific) were calculated with Species Identifier v.1.8 (Meier *et al.*, 2006).

Morphometrics of male genitalia

Genitalia of 45 males were examined, representative for populations ranging from Iberia to southern Russia and Iran (Supporting Information, Table S1). They were prepared according to the following

Downloaded from https://academic.oup.com/zoolinnean/advance-article/doi/10.1093/zoolinnean/zlaa171/6069183 by guest on 08 January 2021

protocol: maceration for 15 min at 95 °C in 10% potassium hydroxide, dissection and cleaning under a stereomicroscope, and storage in tubes with glycerine. Genitalia were immersed in a thin layer of distilled water, pressed under a coverslip and visualized with a Carl Zeiss Stemi 2000-C stereomicroscope equipped with a CMEX PRO-5 DC.5000p digital camera. Genitalia terminology follows Higgins (1975).

A combination of landmarks and sliding semilandmarks (Bookstein, 1997) was applied to the outlines of the cucullus (four landmarks and 20 semilandmarks) and gnathos (three landmarks and seven semilandmarks). We considered as landmarks points that could be identified precisely, whereas the semilandmarks were allowed to slide along the outline trajectory (Fig. 3A). In butterflies, the cucullus and the gnathos are sometimes highly variable among closely related species. Diverse shapes in these structures characterize different species of Hesperiidae, and European and north African species of the genus *Muschampia* typically experience interspecific variation in these structures (Higgins, 1975).

Landmarks were digitized with the program TPSDIG v.2.32 (Rohlf, 2018). In order to remove non-shape variation and to superimpose the objects in a common coordinate system, a generalized Procrustes analysis was applied to the landmark data (Adams *et al.*, 2004). Partial warps were calculated using the shape residuals from the generalized Procrustes analysis. By applying principal components analyses to partial warps, relative warps (RWs; principal components, PCs) were obtained and used as variables in subsequent analyses (Bookstein, 1997). Centroid size, the square root of the sum of squared distances of all the landmarks, has also been calculated as the most appropriate measure for overall size (Bookstein, 1997).

We tested the existence of a signature of diversification in genitalia among the main clades highlighted by molecular markers by using partial least squares discriminant analysis (PLSDA). A series of PLSDAs were carried out by using shape variables (PCs) and the hypothesis for species attribution as the grouping variable. Given that relative warps can be particularly numerous (2× number of landmarks minus four), overfitting had to be avoided. We thus applied a sparse PLSDA (Lê Cao *et al.*, 2011) by including five shape variables in each component. Finally, to evaluate the degree of diversification as a percentage of cases that can be attributed blindly to their group, we applied a jackknife (leave-one-out) algorithm and classified each specimen individually (e.g. Platania *et al.* 2020). These analyses were carried out with the 'splsda' and 'predict' functions in the *mixOmics* R package (Rohart *et al.*, 2017). We compared the centroid size among putative species by using a Kruskal–Wallis test, paired with post hoc tests if cases of an overall significant effect were found.

Comparison of spatial distribution of diversity between markers

The *p*-distance dissimilarity matrices for *COI* and ITS2 were projected in two dimensions by principal coordinates analysis using the 'cmdscale' R function. A two-dimensional representation of the variation in male genitalia was obtained by PLSDA as described above. To facilitate a direct comparison of the patterns of different markers, we eliminated the effect of location and rotation among bidimensional representations with Procrustes analyses, using the *COI* configuration as a reference. Given that different markers were represented by different sets of specimens, we used the 'recluster.procrustes' function of the *recluster* R package (Dapporto *et al.*, 2013), which maximizes similarities among configurations on the basis of partly overlapping data sets (Dapporto *et al.*, 2014). After Procrustes analysis, the aligned bidimensional configurations for specimens were projected in the red–green–blue colour space using the same package. Specimens were grouped when belonging to the same square of 2° × 2° of latitude and longitude, and their individual red–green–blue colours were plotted on a map using pie charts.

Species delimitation

To investigate species limits, we used the software BPP v.4.2.9 (Flouri *et al.*, 2018), running the analysis type 'A10' (Yang & Rannala, 2010; Rannala & Yang, 2013), which implements a reversible-jump Markov chain Monte Carlo (MCMC) species delimitation algorithm. Here, we used only those individuals that had both markers available, which we assigned to species according to Figure 2A. Given that the markers are nuclear and mitochondrial, flags 'heredity' and 'locusrate' were activated. For θ and τ, we explored three scenarios (Leaché & Fujita, 2010): (1) assuming relatively large ancestral population sizes and deep divergences, $\theta \sim IG(3, 0.2)$ and $\tau 0 \sim IG(3, 0.2)$, both with a prior mean = 0.1 and variance = 0.01; (2) assuming relatively small ancestral population sizes and shallow divergences among species, $\theta \sim IG(3, 0.002)$ and $\tau 0 \sim IG(3, 0.002)$, both with a prior mean = 0.001 and variance = 10^{-6}; and (3) assuming large ancestral populations sizes, $\theta \sim IG(3, 0.2)$, and relatively shallow divergences among species, $\tau 0 \sim IG(3, 0.002)$. We ran analyses for 5×10^5 MCMC generations, with a burn-in of 5×10^4 generations and sampling every five generations. For each scenario, we ran BPP twice to check that our results were consistent across runs.

Downloaded from https://academic.oup.com/zoolinnean/advance-article/doi/10.1093/zoolinnean/zlaa171/6069183 by guest on 08 January 2021

We implemented the Bayesian Poisson tree processes (bPTP) model (Zhang *et al.*, 2013) independently for the ITS2 and *COI* markers and using the BEAST phylogeny as input. We selected 500 000 MCMC generations, a thinning value of 100 and a burn-in of 10%.

RESULTS

NUCLEAR DNA

The Bayesian and ML phylogenies based on ITS2 sequences retrieved three clades (Fig. 2A; Supporting Information, Fig. S1) that we defined as three distinct species: (1) *M. proto*, including all Ibero-African and French individuals [posterior probabilities (PP) = 1; bootstrap = 91];(2)*Muschampia alta* (Schwingenschuß, 1942) comb. & stat. nov., consisting of samples from southern Italy and the Balkan Peninsula (PP = 1; bootstrap = 99); and (3) *M. proteides*, encompassing all the populations occurring east of the Balkan Peninsula (PP = 1; bootstrap = 77). We recovered *M. alta* and *M. proteides* as sister taxa, although the support was low (PP and bootstrap < 0.7/70). In the Bayesian phylogeny, *M. proteides* was divided into two divergent clades (PP = 0.99 and PP = 1), one exclusive to the Talysh Mountains (Azerbaijan). We did not detect notable intraspecific structure in the other species. Minimum uncorrected pairwise distances between these species were 2.55% between *M. proto* and *M. alta*, 2.53% between *M. proto* and *M. proteides* and 2.87% between *M. alta* and *M. proteides*. Maximum uncorrected pairwise distances within species were 0.4% in *M. proto*, 0.59% in *M. alta* and 2.22% in *M. proteides*.

MITOCHONDRIAL DNA

The Bayesian and ML phylogenies based on *COI* sequences recovered the *M. proto* complex as monophyletic (PP = 0.78, bootstrap = 82). The complex diverged from *Muschampia mohammed* (Oberthür, 1887) (the sibling species according to Zhang *et al.*, 2020) ~2.3 (95% highest posterior density range = 1.5–3.2)

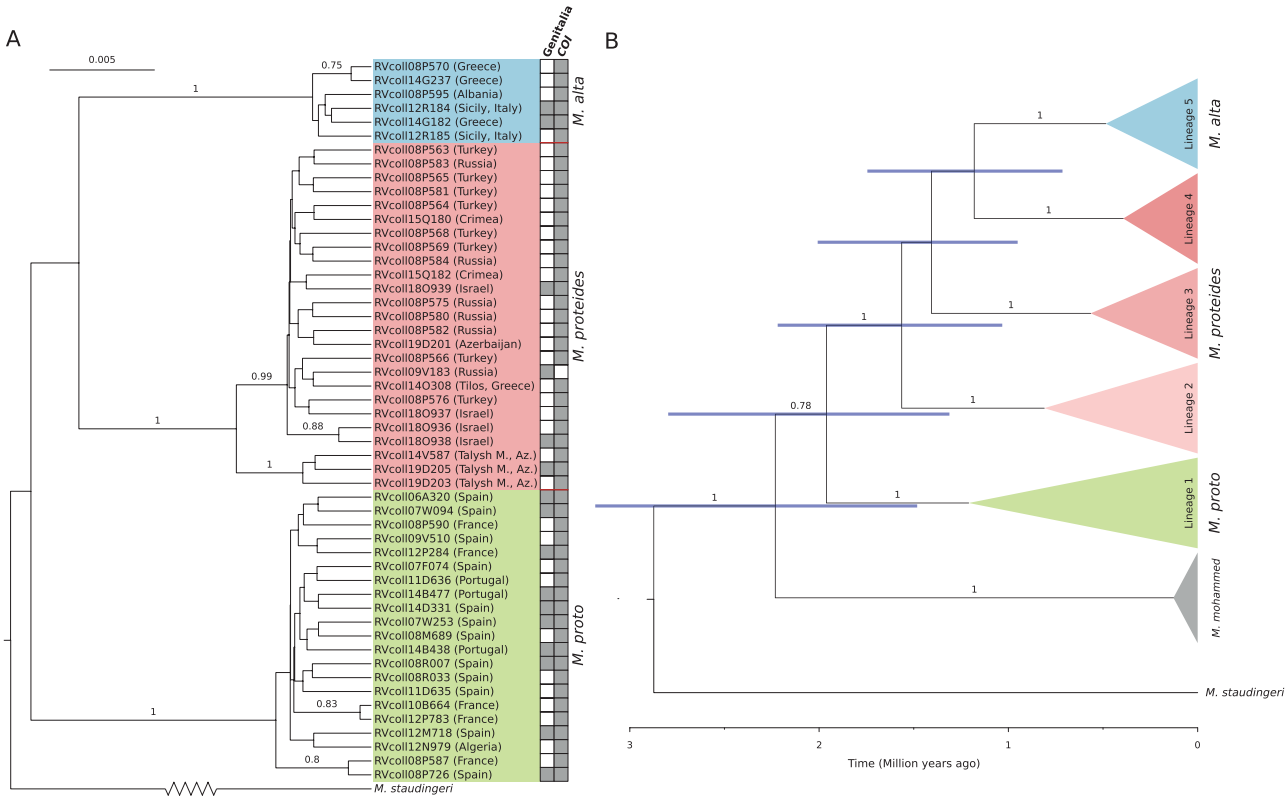

Figure 2. A, phylogenetic tree based on ITS2 data obtained through Bayesian inference. Posterior probabilities > 0.7 are indicated. Scale units are presented in substitutions per site. For each sample, boxes are filled if the genitalia were measured and/or the C*OI* gene was sequenced. B, C*OI* gene tree obtained through Bayesian inference, with the main groups collapsed. The *x*-axis indicates time (in millions of years), and the blue bars show the 95% highest posterior density range for the posterior distribution of node ages.

Downloaded from https://academic.oup.com/zoolinnean/advance-article/doi/10.1093/zoolinnean/zlaa171/6069183 by guest on 08 January 2021

Mya. *Muschampia proto* and *M. alta* were monophyletic and supported, but *M. proteides* was recovered as paraphyletic, although this was not supported (Fig. 2B; Supporting Information, Figs S2, S3). The Bayesian phylogeny highlighted five clades. Among these, lineage 1 (PP = 1; bootstrap = 93), including populations from southern France, Iberia and northern Africa (*M. proto*), was sister to the rest (in agreement with the ITS2 phylogeny), with an estimated age of ~2 (1.3–2.8) Mya. In lineage 1, the Algerian individual displayed a notable divergence (Supporting Information, Figs S2, S3) and a minimum uncorrected pairwise distance of 2.1% compared with the remaining individuals within this clade. The other African specimens, from Morocco, grouped with the European ones in the clade. Lineage 2 (PP = 1; polyphyletic in the ML phylogeny) originated ~1.6 (1.0–2.2) Mya and lineage 3 (PP = 1; bootstrap = 99) ~1.4 (1.0–2.0) Mya. Finally, lineage 4 (PP = 1; bootstrap = 98) and lineage 5 (PP = 1, bootstrap = 94), treated as *M. alta*, split ~1.2 (0.7–1.8) Mya. According to the subset of specimens with ITS2 data available, clades 2, 3 and 4 were attributable to *M. proteides*. In the *COI* tree, the *M. proteides* specimens from the Talysh Mountains (Azerbaijan), which were differentiated based on the nuclear marker (Fig. 2A), formed a recently emerged group inside *M. proteides* lineage 3 (Fig. 2B; Supporting Information, Fig. S2). In *M. alta*, Italian individuals formed a clade (PP = 0.7; bootstrap < 70).

Minimum interspecific genetic distances were 3.79% between *M. proto* and *M. alta*, 2.96% between *M. proto* and *M. proteides* and 2.08% between *M. alta* and *M. proteides*. Maximum uncorrected pairwise distances within species were 2.5% in *M. proto*, 0.93% in *M. alta* and 4.02% in *M. proteides*.

Genitalia

Among the 44 relative warps extracted for the cucullus and 16 relative warps obtained for the gnathos, only four showed absolute values of loading > 0.3 in the PLSDA (cucullus RW1 and RW2; gnathos RW1 and RW3). Being among the first RWs extracted for each process, these variables explained most of the variation in shape of their processes (56.32% for cucullus and 58.47% for gnathos), indicating that most of the variability in shape among specimens was related to differences among the proposed species. Accordingly, a PLSDA plot clearly separated the three groups of specimens (Fig. 3B) that corresponded to *M. proto* (12 genitalia), *M. alta* (16) and *M. proteides* (17). When blindly attributed in a jackknife procedure, 97.2% of samples were correctly classified and only one was misclassified. The centroid size of genitalia did not differ among the three taxa (Kruskal–Wallis $\chi^2 = 3.788$, d.f. = 2, *P*-value = 0.151).

The analysis of thin plate splines allowed inspection of the differences in landmark configurations among average values of the four RWs for the different species. According to the PLSDA results (Fig. 3B), *M. proto* mostly differed from the other two species by the shape of the cucullus, which had a larger dorsal process (Fig. 3C). At the opposite extreme, *M. alta* showed the most reduced process. The gnathos was much shorter in *M. proteides* than in the other two species (Fig. 3C). A representative picture of the genitalia of each species is shown in the Supporting Information (Fig. S4).

Comparison of spatial distribution among markers

A comparison of the spatial distribution of differentiation among markers (Fig. 4A–F) showed a virtually perfect concordance with the three species proposed. The only discrepancy in nuclear and mitochondrial DNA markers was attributable to *M. proteides* samples belonging to mitochondrial lineages 2 and 4 (Fig. 4A–D).

Species delimitation

BPP using both loci separated the complex into the three suggested species with a PP = 1 in the three scenarios studied. Using the ITS2 marker, bPTP recovered the three species in both ML and Bayesian solutions with supports of 0.94 for *M. alta*, 0.88 for *M. proto* and 0.72 for *M. proteides*. In contrast, using the *COI* barcode fragment, bPTP suggested 22 species in the ML solution and 45 in the Bayesian solution; none corresponded to our proposal.

TAXONOMY OF THE *MUSCHAMPIA PROTO* COMPLEX

Muschampia proto (Ochsenheimer, 1808)

Type locality: Portugal.

Original reference: Ochsenheimer F, 1808. *Die Schmetterlinge von Europa. Erster Band. Zweyte Abtheilung*. Leipzig: G. Fleischer.

Original name: Papilio proto.

Monophyletic in the phylogenies based on nuclear and mitochondrial markers. Minimum uncorrected *p*-distances are 2.55% (ITS2) and 3.79% (*COI*) with respect to *M. alta* and 2.53% (ITS2) and 2.96% (*COI*) with respect to *M. proteides*. Morphometric analyses of the genitalia separate *M. proto* from the other species well. It can be distinguished from *M. proteides*

Downloaded from https://academic.oup.com/zoolinnean/advance-article/doi/10.1093/zoolinnean/zlaa171/6069183 by guest on 08 January 2021

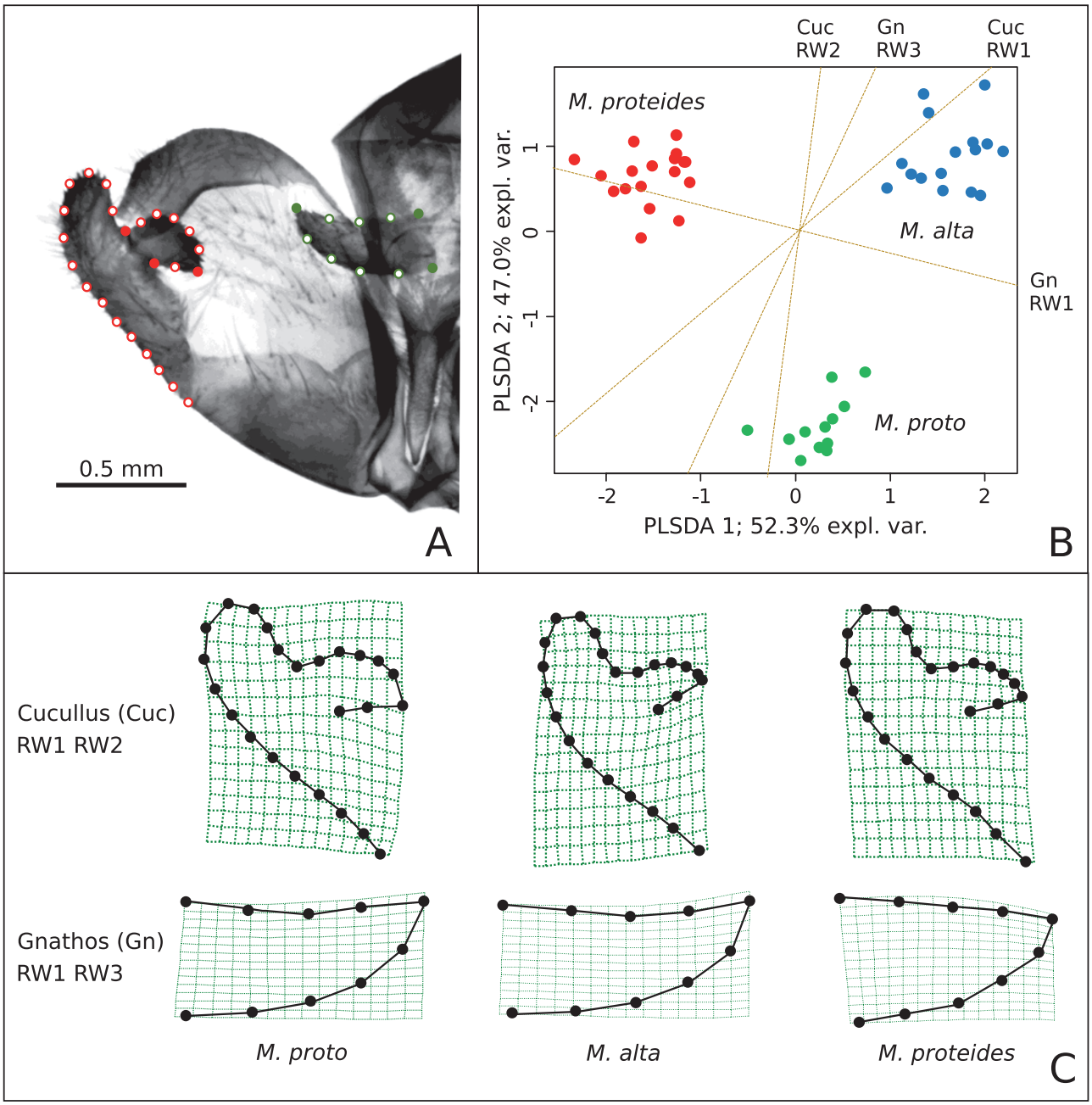

Figure 3. Geometric morphometrics of male genitalia. A, the location of fixed landmarks (filled circles) and sliding semilandmarks (open circles) on the cucullus (red) and gnathos (green). B, the partial least squares discriminant analysis (PLSDA) results, showing specimens of the three species as dots of different colours and the relative warp (RW) scores as dotted lines (Cuc, cucullus; Gn, gnathos). C, thin plate splines representing deformations corresponding to the average values shown by the three species in the relative warps selected by PLSDA as those most involved in the discrimination of the groups.

by the long gnathos, sometimes with the tip rounded. Compared with *M. alta* and *M. proteides*, it has a wider dorsal process of the cucullus. No diagnostic traits in the wings to distinguish it from *M. alta* and *M. proteides* were observed, although no morphometric analyses were performed.

Downloaded from https://academic.oup.com/zoolinnean/advance-article/doi/10.1093/zoolinnean/zlaa171/6069183 by guest on 08 January 2021

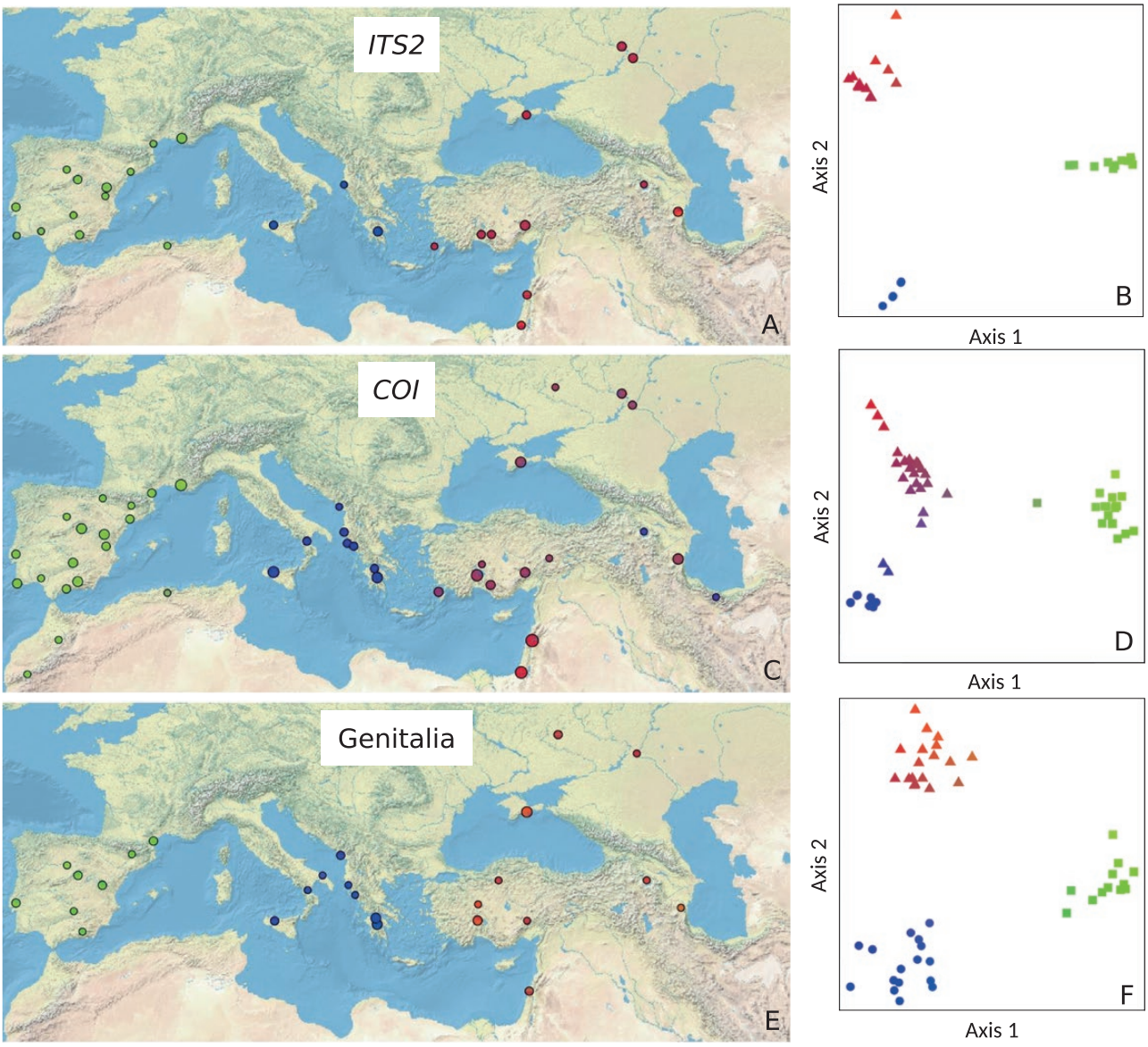

Figure 4. Comparison of spatial distributions of diversity for ITS2, *COI* and morphology of the male genitalia. A, C, E, the specimens have been projected in the red–green–blue colour space, and the resulting colours were plotted in pie charts grouping specimens from the same 2° × 2° latitude–longitude squares (maps on the left). B, D, F, representations of principal coordinates analyses based on dissimilarity matrices for genetic markers and of partial least squares discriminant analysis for male genitalia (circles, *Muschampia alta*; squares, *Muschampia proto*; triangles, *Muschampia proteides*).

Distribution: North-western Africa, the Iberian Peninsula and the Mediterranean coast of France.

MUSCHAMPIA ALTA (SCHWINGENSCHUSS, 1942) COMB. & STAT. NOV.

Type locality: Taormina, Sicily, Italy.

Original reference: Schwingenschuß L. 1942. Eine Falterausbeute aus Sizilien. I. Teil. Macrolepidoptera. *Zeitschrift des Wiener Entomologen-vereines* 27: 177–184.

Original name: Hesperia proto alta.

Monophyletic in the phylogenies based on nuclear and mitochondrial markers, which show that this taxon is sister to *M. proteides*. Minimum uncorrected *p*-distances are 2.55% (ITS2) and 3.79% (*COI*) with respect to *M. proto* and 2.87% (ITS2) and 2.08% (*COI*) with respect to *M. proteides*. It can be distinguished

Downloaded from https://academic.oup.com/zoolinnean/advance-article/doi/10.1093/zoolinnean/zlaa171/6069183 by guest on 08 January 2021

from *M. proteides* by the long gnathos. Morphometrics of the male genitalia highlights differences in the cucullus with respect to *M. proto* and *M. proteides*, in particular a smaller dorsal process of the cucullus. No diagnostic traits in the wings to distinguish it from *M. proto* and *M. proteides* were observed, although no morphometric analyses were performed.

Distribution: Sicily, southern Italian Peninsula and Balkan Peninsula. It is probable that the species is also present in the western Aegean Islands.

MUSCHAMPIA PROTEIDES (WAGNER, 1929)

Type locality: Akşehir, Turkey.

Original reference: Wagner F. 1929. Weiterer Beitrag zur Lepidopteren-Fauna Inner-Anatoliens. *Mitteilungen der Münchner Entomologischen Gesellschaft* 19: 1–28.

Original name: Hesperia proto proteides.

Nuclear and mitochondrial markers displayed differences with respect to *M. proto* and *M. alta*. We recover it as monophyletic in the phylogeny based on ITS2. However, it is paraphyletic with respect to *M. alta* in the *COI* phylogeny. Minimum uncorrected *p*-distances are 2.87% (ITS2) and 2.08% (*COI*) with respect to *M. alta* and 2.53% (ITS2) and 2.96% (*COI*) with respect to *M. proto*. The gnathos of the male genitalia is typically wide and short, a character useful to differentiate it from *M. proto* and *M. alta*. No diagnostic traits in the wings to distinguish it from *M. proto* and *M. alta* were observed, although no morphometric analyses were performed.

Distribution: From eastern Europe (Ukraine, south-western Russia), Anatolia and Israel to the north-eastern coast of the Caspian Sea and Iran. Its distribution further east is unclear. We confirm its presence in the Greek islands of Symi and Tilos, which suggests that it is the only species of the triplet found in the eastern Aegean Islands.

DISCUSSION

THE *MUSCHAMPIA PROTO* COMPLEX, A TRIPLET OF CRYPTIC SPECIES

The results of the nuclear DNA phylogeny (Figs 2A, 4A, B), morphometrics of the male genitalia (Figs 3, 4E, F) and species delimitation analyses (BPP for both loci together and bPTP for ITS2 alone) support the division of the *M. proto* complex into three species with the following geographical distributions (Fig. 1): (1) *M. proto*, present in north Africa, the Iberian Peninsula and southern France; (2) *M. alta*, present in southern Italy and the Balkan Peninsula; and (3) *M. proteides*, distributed east of the Balkan Peninsula. This taxonomic arrangement adds two new species to the European butterfly fauna, which now reaches 498 species (Wiemers *et al.*, 2018). The species proposed can be differentiated based on any of the three traits analysed here, and their characteristics are explained in the '*Taxonomy of the* Muschampia proto *complex*' section above.

The division into three species is clear in the ITS2 phylogeny (Fig. 2A), where we recover each of them as monophyletic with a PP = 1 and bootstrap = 0.77–0.99. Minimum uncorrected *p*-distances between them range from 2.53 to 2.87%, which in this marker can be regarded as usual values for interspecific divergences for Lepidoptera (e.g. Wiemers *et al.*, 2010; Vasudev, 2013). However, in the *COI* phylogeny *M. proteides* is paraphyletic with respect to *M. alta*, although this pattern is not well supported in any analysis. *Muschampia proteides* and *M. alta* also have the lowest minimum distance for *COI* (2.08%) compared with those found between *M. proteides* and *M. proto* (2.96%) and between *M. alta* and *M. proto* (3.79%). Although this divergence is not very high, similar or lower values are sometimes found between butterfly species (Ashfaq *et al.*, 2013; Dincă *et al.* 2013; Huemer *et al.*, 2014), and species that share the same barcode sequences are not exceptional (Dincă *et al.*, 2015).

The phylogeny inferred from *COI* (Fig. 2B; Supporting Information, Figs S2, S3) displays lineages that are absent in the ITS2 tree (Fig. 2A). This is not uncommon and might be the result of the higher mutation rate of mitochondrial DNA (Allio *et al.*, 2017) and/or of processes of isolation followed by panmixia (Hinojosa *et al.*, 2019; Pazhenkova & Lukhtanov, 2019). For example, in *M. alta* the mitochondrial gene slightly separates the Balkan specimens from the Italian ones in two clades (with a PP < 0.7 for the Balkan and PP = 0.7 for the Italian; both bootstrap values are < 70), a division that is not reflected in the nuclear marker. These results would point to a case of incomplete lineage sorting in the nuclear marker. In *M. proto*, a sample from Algeria has a considerable minimum distance of 2.1% in the barcode compared with the rest, but it maintains an almost identical ITS2 sequence to the other conspecifics. *COI* haplotypes are shared between Moroccan and Iberian specimens, which indicates that gene flow between Iberia and Africa occurred recently. *Muschampia proteides* is divided into two clades in the ITS2 phylogeny, but forms three clades in the *COI* phylogeny; in fact, it is the most variable species in both markers, with a maximum intraspecific distance that reaches 2.22%

Downloaded from https://academic.oup.com/zoolinnean/advance-article/doi/10.1093/zoolinnean/zlaa171/6069183 by guest on 08 January 2021

in the ITS2 and 4.02% in the *COI*. One of the ITS2 lineages is formed exclusively by specimens from the Talysh Mountains (Azerbaijan), whereas the other Azerbaijani individual present in the ITS2 phylogeny, from the Lesser Caucasus, is grouped with the rest of *M. proteides*. This lineage does not correspond to any of the three main mitochondrial lineages and is retrieved as a relatively recent split inside mitochondrial lineage 2 (Supporting Information, Figs S2, S3). The analysis of male genitalia does not place them as outliers of *M. proteides*, and no distinctive wing patterns have been observed; therefore, we consider the population from the Talysh Mountains as *M. proteides*. We hypothesize that alternate processes of isolation and panmixia generated the observed genetic pattern.

The existence of three entities is reinforced by the measurements of male genitalia. They clearly distinguish three groups that exhibit notable differences (Figs 3, 4E, F). All but one of the specimens (45 out of 46) were correctly classified when blindly attributed in a jackknife procedure. More precisely, species can be distinguished by the relative dimension of the dorsal process of the cucullus compared with the tip of the valva (for details, see the section '*Taxonomy of the* Muschampia proto *complex*'). *Muschampia proteides* displays the most obvious differences in genitalia because of its short and wide gnathos, a distinctive trait also demonstated by Kemal & Koçak (2017).

Before the present study, the distribution of *M. proteides* was considered by some authors to be restricted to Anatolia, Transcaucasia, the Near East and Iran, whereas populations occurring in the northern part of the Black and Caspian Seas were considered to belong to *M. proto* (Tshikolovets, 2011). We demonstrate here that there are no distinctive traits between these populations and that they all belong to *M. proteides*. Additionally, *M. proto* and *M. proteides* have been cited in sympatry in Anatolia and the Levant (Hesselbarth *et al.*, 1995; Benyamini, 2010; Tshikolovets, 2011). Hesselbarth *et al.* (1995) even reported differences in wing pattern and coloration to distinguish both species in sympatry. We analysed individuals from this region corresponding to the two forms (Supporting Information, Fig. S5) described by Hesselbarth *et al.* (1995) and we did not find differences in the molecular markers (all were *M. proteides* belonging to *COI* lineage 2 apart from one belonging to lineage 3). Thus, neither *M. proto* nor *M. alta* seems to be present in Anatolia, where apparently only *M. proteides* occurs.

Phylogeographical history of the complex

The hypothetical phylogeographical history of the complex that can be reconstructed based on the molecular evidence, with main events dated according to the mitochondrial chronogram (Fig. 2B; Supporting Information, Figs S2, S3), includes the following main steps:

1. Approximately 2.3 (1.5–3.2) Mya, the ancestor of the *M. proto* complex and *M. mohammed* split, with the former probably remaining in Europe and Asia and the latter in northern Africa.
2. Approximately 2.0 (1.3–2.8) Mya, Iberian populations separated from the eastern ones. A division within this clade occurred ~1.3 (0.7–1.9) Mya, when African and Iberian populations split, probably as a result of the colonization of Africa. Gene flow from Iberia to Morocco took place again recently.
3. The eastern lineage (nowadays *M. proteides* and *M. alta*) possibly experienced periods of fragmentation and isolation that drove mitochondrial differentiation, but it would appear that populations subsequently admixed. Isolation events happened ~1.6 (1.0–2.2) Mya, generating lineage 2, and ~1.4 (1.0–2.0) Mya, resulting in lineage 3.
4. At the latest 1.2 (0.7–1.8) Mya, *M. alta* and *M. proteides* split. We did not find evidence of posterior gene flow between them, although nowadays they are geographically close in the Aegean Islands. More recently, ~0.5 (0.3–0.8) Mya, the Italian and Balkan populations of *M. alta* diverged.

Conclusions

We conclude that *M. proto s.l.* is a complex of three cryptic species that diverged between ~2.0 and 1.2 Mya: *M. proto*, present in north Africa, the Iberian Peninsula and southern France; *M. alta*, found in southern Italy and the Balkan Peninsula; and *M. proteides*, which is distributed east of the Balkan Peninsula. All three are well differentiated and monophyletic in both mitochondrial and nuclear phylogenies except for *M. proteides*, which, although diverged, is paraphyletic with low support with respect to *M. alta* in the *COI* tree. Most species delimitation analyses performed support the three species hypothesis. Morphometrics of the male genitalia reveals that *M. proteides* has a shorter and wider gnathos compared with *M. alta* and *M. proto*, and all species show a different relative dimension of the dorsal process of cucullus compared with the tip of the valva. This research adds two new species to the European butterfly fauna and shows that hidden diversity still awaits discovery even in taxa and regions that are generally well known.

ACKNOWLEDGEMENTS

We are grateful to all the colleagues who provided samples used in this study. Financial support for

Downloaded from https://academic.oup.com/zoolinnean/advance-article/doi/10.1093/zoolinnean/zlaa171/6069183 by guest on 08 January 2021

this research was provided by projects PID2019-107078GB-I00 / Agencia Estatal de Investigación / 10.13039/501100011033 and 2017-SGR-991 (Generalitat de Catalunya) to R.V., by a predoctoral fellowship BES-2017-080641 (Ministerio de Economía y Empresa de España) to J.C.H., by the Academy of Finland (decision no. 328895) to V.D. and by project RSF 19-14-00202 (from the Russian Science Foundation to the Zoological Institute of the Russian Academy of Sciences) to V.A.L., and by the project Servizio di attuazione delle azioni per la protezione degli impollinatori e diffusione dell'entomofauna del Parco Nazionale dell'Alta Murgia to L.D.

REFERENCES

Adams DC, **Rohlf FJ**, **Slice DE. 2004.** Geometric morphometrics: ten years of progress following the 'revolution'. *Italian Journal of Zoology* **71:** 5–16.

Alizon S, **Kucera M**, **Jansen VA. 2008.** Competition between cryptic species explains variations in rates of lineage evolution. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 12382–12386.

Allio R, **Donega S**, **Galtier N**, **Nabholz B. 2017.** Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Molecular Biology and Evolution* **34:** 2762–2772.

Ashfaq M, **Akhtar S**, **Khan AM**, **Adamowicz SJ**, **Hebert PD. 2013.** DNA barcode analysis of butterfly species from Pakistan points towards regional endemism. *Molecular Ecology Resources* **13:** 832–843.

Benyamini D. 2010. *A field guide to the butterflies of Israel*. Jerusalem: Keter Publishing House.

Bickford D, **Lohman DJ**, **Sodhi NS**, **Ng PK**, **Meier R**, **Winker K**, **Ingram KK**, **Das I. 2007.** Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution* **22:** 148–155.

Bookstein FL. 1997. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical Image Analysis* **1:** 225–243.

Bouckaert R, **Heled J**, **Kühnert D**, **Vaughan T**, **Wu CH**, **Xie D**, **Suchard MA**, **Rambaut A**, **Drummond AJ. 2014.** BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10:** e1003537.

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 of the United States of America* **91:** 6491–6495.

Ceballos G, **Ehrlich PR. 2009.** Discoveries of new mammal species and their implications for conservation and ecosystem services. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 3841–3846.

Dapporto L, **Cini A**, **Menchetti M**, **Vodă R**, **Bonelli S**, **Casacci LP**, **Dincă V**, **Scalercio S**, **Hinojosa JC**, **Biermann H**, **Forbicioni L**, **Mazzantini U**, **Venturi L**, **Zanichelli F**, **Balletto E**, **Shreeve TG**, **Dennis RLH**, **Vila R. 2017.** Rise and fall of island butterfly diversity: understanding genetic differentiation and extinction in a highly diverse archipelago. *Diversity and Distributions* **23:** 1169–1181.

Dapporto L, **Ramazzotti M**, **Fattorini S**, **Talavera G**, **Vila R**, **Dennis RLH. 2013.** Recluster: an unbiased clustering procedure for beta-diversity turnover. *Ecography* **36:** 1070–1075.

Dapporto L, **Vodă R**, **Dincă V**, **Vila R. 2014.** 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.

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

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, **Montagud S**, **Talavera G**, **Hernández-Roldán J**, **Munguira ML**, **García-Barros E**, **Hebert PD**, **Vila R. 2015.** DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. *Scientific Reports* **5:** 12395.

Dincă V, **Wiklund C**, **Lukhtanov VA**, **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.

Esteban GF, **Finlay BJ. 2010.** Conservation work is incomplete without cryptic biodiversity. *Nature* **463:** 293.

Flouri T, **Jiao X**, **Rannala B**, **Yang Z. 2018.** Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Molecular Biology and Evolution* **35:** 2585–2593.

García-Barros E, **Munguira ML**, **Stefanescu C**, **Vives Moreno A. 2013.** *Fauna Ibérica, Vol. 37. Lepidoptera. Papilionoidea*. Madrid: Museo Nacional de Ciencias Naturales & Consejo Superior de Investigaciones Científicas.

Hebert PD, **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 of the United States of America* **101:** 14812–14817.

Hernández-Roldán JL, **Dapporto L**, **Dincă V**, **Vicente JC**, **Hornett EA**, **Šíchová J**, **Lukhtanov VA**, **Talavera G**, **Vila R. 2016.** Integrative analyses unveil speciation linked to host plant shift in *Spialia* butterflies. *Molecular Ecology* **25:** 4267–4284.

Hesselbarth G, **Van Oorschot H**, **Wagener S. 1995.** *Die Tagfalter der Turkei*. Bocholt: Goecke & Evers.

Downloaded from https://academic.oup.com/zoolinnean/advance-article/doi/10.1093/zoolinnean/zlaa171/6069183 by guest on 08 January 2021

Higgins LG. 1975. *The classification of European butterflies*. London: Harper Collins.

Hinojosa JC, **Koubínová D**, **Szenteczki MA**, **Pitteloud C**, **Dincă V**, **Alvarez N**, **Vila R. 2019.** A mirage of cryptic species: Genomics uncover striking mitonuclear discordance in the butterfly *Thymelicus sylvestris*. *Molecular Ecology* **28:** 3857–3868.

Huemer P, **Mutanen M**, **Sefc KM**, **Hebert PD. 2014.** Testing DNA barcode performance in 1000 species of European lepidoptera: large geographic distances have small genetic impacts. *PLoS One* **9:** e115774.

Kearse M, **Moir R**, **Wilson A**, **Stones-Havas S**, **Cheung M**, **Sturrock S**, **Buxton S**, **Cooper A**, **Markowitz S**, **Duran C**, **Thierer T**, **Ashton B**, **Meintjes P**, **Drummond A. 2012.** Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28:** 1647–1649.

Kemal M, **Koçak AO. 2017.** Notes on *Muschampia proteides* (Wagner) (Lepidoptera, Hesperiidae). *Cesa News* **140:** 1–6.

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.

Leaché AD, **Fujita MK. 2010.** Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society B: Biological Sciences* **277:** 3071–3077.

Lukhtanov VA, **Dantchenko AV. 2017.** A new butterfly species from south Russia revealed through chromosomal and molecular analysis of the *Polyommatus* (*Agrodiaetus*) *damonides* complex (Lepidoptera, Lycaenidae). *Comparative Cytogenetics* **11:** 769–795.

Lumley LM, **Sperling FA. 2010.** Integrating morphology and mitochondrial DNA for species delimitation within the spruce budworm (*Choristoneura fumiferana*) cryptic species complex (Lepidoptera: Tortricidae). *Systematic Entomology* **35:** 416–428.

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

Miller MA, **Schwartz T**, **Pickett BE**, **He S**, **Klem EB**, **Scheuermann RH**, **Passarotti M**, **Kaufman S**, **O'Leary MA. 2015.** A RESTful API for access to phylogenetic tools via the CIPRES Science Gateway. *Evolutionary Bioinformatics* **11:** 43–48.

Pazhenkova EA, **Lukhtanov VA. 2019.** Nuclear genes (but not mitochondrial DNA barcodes) reveal real species: evidence from the *Brenthis* fritillary butterflies (Lepidoptera, Nymphalidae). *Journal of Zoological Systematics and Evolutionary Research* **57:** 298–313.

Pfenninger M, **Schwenk K. 2007.** Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology* **7:** 121.

Platania L, **Vodă R**, **Dincă V**, **Talavera G**, **Vila R**, **Dapporto L. 2020.** Integrative analyses on Western Palearctic *Lasiommata* reveal a mosaic of nascent butterfly species. *Journal of Zoological Systematics and Evolutionary Research* **58:** 809–822.

Quek SP, **Davies SJ**, **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; international journal of organic evolution* **58:** 554–570.

Rambaut A, **Drummond AJ**, **Xie D**, **Baele G**, **Suchard MA. 2018.** Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **67:** 901–904.

Rannala B, **Yang Z. 2013.** Improved reversible jump algorithms for Bayesian species delimitation. *Genetics* **194:** 245–253.

Rohart F, **Gautier B**, **Singh A**, **Lê Cao KA. 2017.** mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Computational Biology* **13:** e1005752.

Rohlf FJ. 2018. *tpsDig32*. Available at: http://life.bio.sunysb.edu/morph/

Scalercio S, **Cini A**, **Menchetti M**, **Vodă R**, **Bonelli S**, **Bordoni A**, **Casacci PL**, **Dincă V**, **Balletto E**, **Vila R**, **Dapporto L. 2020.** How long is 3 km for a butterfly? Ecological constraints and functional traits explain high mitochondrial genetic diversity between Sicily and the Italian Peninsula. *Journal of Animal Ecology* **89:** 2013–2026.

Schmitt T, **Louy D**, **Zimmermann E**, **Habel JC. 2016.** Species radiation in the Alps: multiple range shifts caused diversification in ringlet butterflies in the European high mountains. *Organisms Diversity & Evolution* **16:** 791–808.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30:** 1312–1313.

Struck TH, **Feder JL**, **Bendiksby M**, **Birkeland S**, **Cerca J**, **Gusarov VI**, **Kistenich S**, **Larsson KH**, **Liow LH**, **Nowak MD**, **Stedje B**, **Bachmann L**, **Dimitrov D. 2018.** Finding evolutionary processes hidden in cryptic species. *Trends in Ecology & Evolution* **33:** 153–163.

Tshikolovets VV. 2011. *Butterflies of Europe and Mediterranean area*. Pardubice: Tshikolovets Publications.

Tuzov VK. 1997. *Guide to the butterflies of Russia and adjacent territories*. Sofia: Pensoft.

Vasudev K. 2013. Molecular diversity of *Conogethes* spp. Guenèe, (Lepidoptera: Crambidae) infesting castor and cardamom. Unpublished D. Phil. Thesis, University of Agricultural Sciences, Bangalore.

Verovnik R, **Wiemers M. 2016.** Species delimitation in the grayling genus *Pseudochazara* (Lepidoptera, Nymphalidae, Satyrinae) supported by DNA barcodes. *ZooKeys* **600:** 131–154.

Vodă R, **Dapporto L**, **Dincă V**, **Vila R. 2015.** Cryptic matters: overlooked species generate most butterfly beta-diversity. *Ecography* **38:** 405–409.

deWaard JR, **Ivanova NV**, **Hajibabaei M**, **Hebert PDN. 2008.** Assembling DNA barcodes: analytical protocols. In: Cristofre M, ed. *Methods in molecular biology: environmental genetics*. Totowa: Humana Press, 275–293.

Wiemers M, **Balletto E**, **Dincă V**, **Fric ZF**, **Lamas G**, **Lukhtanov V**, **Munguira ML**, **Van Swaay CAM**, **Vila R**,

Downloaded from https://academic.oup.com/zoolinnean/advance-article/doi/10.1093/zoolinnean/zlaa171/6069183 by guest on 08 January 2021

Vliegenthart A, **Wahlberg N**, **Verovnik R. 2018.** An updated checklist of the European Butterflies (Lepidoptera, Papilionoidea). *ZooKeys* **811:** 9–45.

Wiemers M, **Stradomsky BV**, **Vodolazhsky DI. 2010.** A molecular phylogeny of *Polyommatus* s. str. and *Plebicula* based on mitochondrial *COI* and nuclear *ITS2* sequences (Lepidoptera: Lycaenidae). *European Journal of Entomology* **107:** 325–336.

Yang Z, **Rannala B. 2010.** Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 9264–9269.

Zhang J, **Brockmann E**, **Cong Q**, **Shen J**, **Grishin NV. 2020.** A genomic perspective on the taxonomy of the subtribe Carcharodina (Lepidoptera: Hesperiidae: Carcharodini). *Zootaxa* **4748:** 182–194.

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**, **Chelazzi G**, **Bonelli S**, **Balletto E**, **Ciofi C. 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.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. ITS2 gene tree obtained through maximum likelihood inference. Bootstrap values > 70 are indicated.
Figure S2. *COI* gene tree obtained through Bayesian inference. Posterior probabilities > 0.7 are indicated. The *x*-axis indicates time (in millions of years before present), and the blue bars show the 95% highest posterior density range for the age posterior distribution of nodes older than 500 000 years.
Figure S3. *COI* gene tree obtained through maximum likelihood inference. Bootstrap values > 70 are indicated.
Figure S4. Comparison of male genitalia of the three species displaying the typical forms. In *Muschampia proto*, the tip of the gnathos is also commonly pointed.
Figure S5. Selection of four *Muschampia proteides* individuals from Anatolia used in this study. Following the morphological criteria of Hesselbarth *et al.* (1995), the first two (RVcoll08P581 and RVcoll08P597) would correspond to the form classified as '*M. proteides*' and the other two (RVcoll08P568 and RVcol08P599) to '*M. proto*'.
Table S1. Specimens used in this study and GenBank accession numbers.

Downloaded from https://academic.oup.com/zoolinnean/advance-article/doi/10.1093/zoolinnean/zlaa171/6069183 by guest on 08 January 2021