



OPEN

Next generation sequencing-aided comprehensive geographic coverage sheds light on the status of rare and extinct populations of *Aporia* butterflies (Lepidoptera: Pieridae)

Valentina Todisco^{1✉}, Raluca Vodă², Sean W. J. Prosser³ & Vazrick Nazari⁴

The Black-veined White *Aporia crataegi* (Linnaeus, 1758), a common and widespread butterfly ranging from northwestern Africa to Europe and Asia, has been extinct in Britain since the 1920s and is on a steady decline in several other parts of its range. In order to investigate genetic diversity within *A. crataegi* and its correspondence with current subspecies-level taxonomy, we barcoded 173 specimens from across its range including, for the first time, extinct populations from Britain and Korea. Using next generation sequencing we also obtained a sequence for *Aporia joubini*, a peculiar taxon from China known only by its type specimen collected in the early twentieth century. Our phylogenetic analysis placed *A. joubini* sister to *A. oberthuri*, although further taxon sampling may reveal a different scheme. Within *A. crataegi*, we observed a shallow and weak mitogenomic structure with only a few distinct lineages in North Africa, Sicily, Iran, and Japan. Eurasian populations, including those extinct in Britain and Korea, clustered into a large set of closely allied lineages, consistent with a recent expansion during the Late Pleistocene glacial period. This study highlights the importance of museum collections and the unique opportunities they provide in documenting species diversity and helping conservation efforts.

Human activities threaten considerable numbers of species and natural populations that, in turn, significantly affect ecosystem functioning^{1,2}. Each species contains immense amounts of genetic information within its geographical populations, and this intra-specific genetic diversity not only is a repository of past geological, climatic, and environmental information, but also influences population persistence and evolutionary potential³. Thus, the resulting permanent loss of genetic information due to local extinctions not only affects evolutionary potential under changing environmental conditions, but also impedes a full understanding of the evolutionary history of species. Natural history collections have long been used by morphologists and taxonomists to probe the evolutionary process and describe biological diversity. These biological archives also offer great opportunities for genetic research in taxonomy, conservation, systematics, and population biology and they represent an invaluable resource for studying extinct taxa.

Next Generation Sequencing (NGS) has revolutionized nearly every field of genetics, but few have profited from it as much as ancient DNA research. NGS has increased the amount of DNA sequence data available from extinct organisms by several orders of magnitude and is extensively used in phylogenetic studies using museum specimens [e.g.^{4,5}].

The Black-veined White, *Aporia crataegi*, is a well-known butterfly widespread in orchards and bushy places in the Palearctic region, from north-western Africa to Japan, across Europe and Asia, although very localized and uncommon in some areas^{6–8}. Its larvae feed on Rosaceae such as *Prunus*, *Pyrus*, and especially *Crataegus*.

¹Unaffiliated, Marino (Rome), Italy. ²Unaffiliated, Torino, Italy. ³Unaffiliated, Guelph, ON N1E 7L4, Canada. ⁴Unaffiliated, Ottawa, ON K2M 2Y1, Canada. ✉email: valentina.todisco@gmail.com

Despite a very simple and uniform wing pattern, numerous local subspecies have been described under *A. crataegi* across its range, and the status of many remain doubtful: Della Bruna et al.⁹ listed 25 subspecies under *A. crataegi*, while Savela¹⁰ recognized 23. The Black-veined White belongs to a diverse genus composed of numerous other species, most of which are confined to southwest China and the Himalaya-Hengduan mountain regions. The phylogenetic position of *Aporia* within Pieridae has been well established^{11,12}. However, its monophyly has been recently questioned by Ding and Zhang¹³, who found that the monotypic genus *Mesapia* Gray, 1856 (type species: *Pieris peloria* Hewitson, [1853]), proposed either as a subgenus of *Aporia*^{14,15} or as a distinct genus^{9,16,17}, is nested within *Aporia* sensu lato (see also¹⁸). This may highlight a need to synonymize *Mesapia* under *Aporia*. Moreover, Ding and Zhang^{13,18} found that another subgenus within *Aporia*, subgn. *Metaporia* (Butler 1870, type species: *Pieris agathon* Gray, 1831) which includes species with a bifid apex of the uncus in male genitalia sensu Della Bruna et al.⁹, is also paraphyletic, and therefore subdivisions of *Aporia* into separate subgenera cannot be supported. Two studies have so far investigated the phylogeny and evolution of the species within this genus, using either a mitochondrial COI fragment alone²⁰ or accompanied by a second nuclear gene¹⁸, although taxon sampling in both of these studies was incomplete. *Aporia crataegi* was once a fairly widespread and locally plentiful butterfly in southern England and Wales. By 1872 it had disappeared from Kent, and according to Allan²¹, the last authentic British record as a native species was a single specimen caught in New Forest in 1880, although it is more widely accepted that the last British specimens were those from Herne Bay in Kent caught during the 1920s. The reasons behind the extinction of *A. crataegi* in Britain remain unknown, and attempts to reintroduce it by releasing hundreds of continental specimens in Kent and elsewhere in Britain failed to re-establish *A. crataegi* on a permanent basis. The populations of *A. crataegi* can fluctuate greatly in numbers over time, but the reasons for this are not fully understood. Major endangerment factors for their decline include habitat destruction, climatic change, parasitism, bird predation, and the use of pesticides (especially herbicides) in orchards^{22,23}. In recent years, strong declines in distribution, or population size of more than 30%, have been recorded for *A. crataegi* from Austria, Romania, Ukraine, Albania, France, Latvia, Norway, Serbia, Algeria, and Morocco²⁴, and the butterfly is also reported extinct from Netherlands, Luxembourg, Czech Republic²⁵, and Korea (since 1997²³). In this study we analysed the mitochondrial COI barcode region for 23 out of 25 subspecies of *A. crataegi* as recognized by Della Bruna et al.⁹ from the Palearctic region, including, for the first time, the extinct British populations. We also obtained a COI barcode from the only known specimen of *A. joubini*, collected in 1902 from China and now hosted in the Natural History Museum of London (NHMUK type). We aimed to: (i) investigate the phylogenetic position of *A. joubini* within the *Aporia* genus, (ii) assess the phylogeographic patterns within *A. crataegi*, (iii) investigate if the current taxonomic entities correspond with these patterns, and (iv) evaluate a continental vs. an autochthon origin of the British specimens.

Materials and methods

Taxon sampling. We examined 173 *A. crataegi* individuals representing 23 subspecies out of the 25 recognized by Della Bruna et al.⁹ from sites across Eurasia and North Africa (Maghreb region) (Fig. 2A and Supplementary Dataset File). We were unable to obtain sequences from ssp. *fert* (Rhodes Island, Greece) and *sachalinensis* (Sakhalin, Russia). We analysed ten British individuals: four collected between 1860 and 1870, before the extinction of the native populations, and six at the beginning of the twentieth century, supposedly re-introduced from the continent. In addition, we analysed four specimens from Iranian populations of *A. leucodice*, and the only known (type) specimen of *A. joubini*, collected in 1902 from China.

DNA extraction. A single leg was removed from each specimen and sent to the Centre for Biodiversity Genomics in Guelph, Canada for DNA extraction, amplification, and sequencing. DNA was extracted from a single leg on a Biomek FX liquid handling robot using a semi-automated DNA extraction protocol²⁶ with glass fiber plates (PALL Acroprep 96 with 3 µm GF membrane over 0.2 µm Bioinert membrane). DNA was eluted in 30–40 µl of ddH₂O pre-warmed to 56 °C and stored at – 80 °C.

Next-generation and Sanger sequencing. Due to DNA degradation in the very old specimens of *A. joubini* and the British *A. crataegi*, the 658 bp COI barcode region was recovered using an NGS-based protocol²⁷. In brief, for each sample, multiple short, overlapping amplicons were generated using nested, multiplex PCR. In order to associate reads with their source specimen, the amplicons were tailed with sample-specific universal molecular identifiers (UMI) before being pooled for sequencing on an Ion Torrent PGM. The short sequence reads were attributed to a sample via the UMIs, and filtered for quality and length before being assembled into a single barcode sequence.

For the remaining (younger) specimens, PCR and Sanger sequencing were carried out following standard procedures for Lepidoptera^{28,29}. Briefly, the 658 bp barcode region was amplified using the primers LepF1 (5' ATTCAACCAATCATAAAGATATTGG 3') and LepR1 (5' TAAACTTCTGGATGTCCAAAAAATCA 3') and sequenced on an ABI 3730XL (Applied Biosystems capillary sequencer). The trace files were edited using CodonCode Aligner 6.0.2 (CodonCode Corporation, Dedham, Massachusetts) and all resulting mtDNA sequences were aligned using the same program. All sequences were submitted to GenBank (Accession Numbers in Supplementary Dataset File) and BOLD system repository (dataset name: DS—APORIA; <https://doi.org/10.5883/DS-APORIA>).

Data set compilation and tree reconstruction procedures. To assess the phylogeographic patterns within *A. crataegi* and the origin of the British specimens, we complemented the 173 specimens sequenced specifically for this study with an additional 24 sequences of *A. crataegi* from public DNA barcoding projects (see Supplementary Dataset File). The phylogenetic position of *A. joubini* within the genus *Aporia* was further

Group	N	H	h	F _s	P	R ₂	P	τ	τ (5%)	τ (95%)	t (ka)	t (5%)	t (95%)
Eurasian haplogroup	114	24	0.65 ± 0.04	-29.24	0	0.095	0	1.011	0.746	1.492	67	49	99
Mediterranean haplogroup	35	12	0.75 ± 0.07	-6.99	0	0.12	0	1.748	0.869	2.895	116	57	191

Table 1. Results of the tests of demographic equilibrium and mismatch analysis in mtDNA phylogeographic groups. N, number of mtDNA sequences; H, number of unique mtDNA haplotypes; h, mtDNA haplotype diversity (\pm SD). F_s, Fu's FS statistic; R₂, Rozas and Ramon-Onsins' R₂ statistic; τ, sudden demographic expansion parameter, with 5% and 95% confidence limits; t, true time since population expansion, from $\tau = 2 \mu T$ (where μ is the substitution rate per gene). Significantly small values of F_s and R₂ are indicated in bold.

assessed, reconstructing the phylogenetic relationships in the subtribe Aporiina. Seventeen GenBank sequences (n = 17) from 12 species of *Aporia* and 6 outgroups, mainly selected from Ding and Zhang¹⁸, were thus added (see Supplementary Dataset File). We excluded *A. martinetti* (KU921265), *A. genestieri* (KU921259) and *Mesapia peloria* (KU921272) sequences from Ding and Zhang¹⁸ due to the questionable quality of the COI sequences in GenBank. All sequences were aligned using CodonCode Aligner 6.0.2. Haplotype and nucleotide diversity were calculated using DnaSP 5.0³⁰. The mean p-distance between groups and its variance (bootstrap method, 500 replicates) was calculated using MEGA 10³¹. To represent the genealogical relationships among mtDNA haplotypes, we calculated a Median Joining (MJ) network using NETWORK 5³². *Aporia* relationships were assessed via Maximum Likelihood (ML). The software package ModelFinder³³ in IQ-TREE 1.6.11³⁴ was used to select the best fitting model of evolution by Bayesian information criterion scores (BIC). The ML analysis was carried out with RaxML 7.0.3³⁵, on the T-REX (Tree and reticulogram REConstruction) web server^(36; https://www.trex.uqam.ca), with the rapid hill-climbing algorithm and 1,000 non-parametric bootstrap inferences. The highest bootstrap values were obtained setting the model of evolution to GTRCAT, searching for the best-scoring ML tree in a single run. The best ML tree was visualized and edited with FigTree version 1.4³⁷.

Tests of demographic equilibrium and population expansion in mtDNA haplogroup. To test demographic equilibrium or events of recent expansion in *A. crataegi*, different sets of mtDNA sequences were selected (Table 1) considering the results of the previous phylogenetic analysis. To detect departures of DNA sequence variability from the expectations of the neutral theory of evolution³⁸, R₂³⁹ and Fu's FS⁴⁰ parameters were calculated. Both have been shown to provide sensitive signals of historical population expansions⁴⁰. While the latter is based on the haplotype distribution and it is expected to be strongly affected by recombination, R₂ is based on mutations and likely to be conservative on recombining regions⁴¹. Moreover, Fu's simulations suggest that FS is a more sensitive indicator of population expansion than other parameters, therefore FS should be regarded as significant if $P < 0.02$ ⁴². Arlequin 3.5⁴³ and DnaSP 5.0³⁰ were used to compute FS ($P < 0.02$) and R₂ ($P < 0.05$) respectively, and to test their statistical significance by simulating random samples (10,000 replicates) under the null hypothesis of selective neutrality and constant population size using coalescent algorithms (both modified from Hudson⁴⁴). Expected mismatched distributions and parameters of sudden expansion $\tau = 2 \mu t$ (τ , sudden demographic expansion parameter; t, true time since population expansion; μ is the substitution rate per gene) were calculated using Arlequin 3.5 by a generalized least-squares approach⁴⁵, under models of pure demographic expansion and spatial expansion^{46,47}. The probability of the data under the given model was assessed by the goodness-of-fit test implemented in Arlequin 3.5. Parameter confidence limits were calculated in Arlequin 3.5 through a parametric bootstrap (1,000 simulated random samples).

Specimens collection statement. We declare that all samples analysed in this study were collected complying with institutional, national, and international guidelines.

Results

Sequencing results. A COI barcode sequence of 552 bp for *A. joubini* and sequences of various lengths (see Supplementary Dataset File) for seven out of ten British *A. crataegi* were recovered with NGS. These sequences perfectly matched with other *Aporia* barcodes, when using BOLD Identification System (IDS). Further confirmation of their validity was provided by the fact that they grouped with sequences from closely allied taxa. The missing data in NGS sequences (see Supplementary Dataset File) as well as the lack of full-length (658 bp) barcode sequences for three British *A. crataegi* samples can be attributed to amplification failure of the COI gene due to DNA degradation. Despite a match of 97% to the other *Aporia*, one British *A. crataegi* sequence was excluded due to branching far from the other *A. crataegi* sequences (its position in the tree was likely an artifact produced by its short sequence length: 130 bp containing two N's). Sanger sequencing produced full-length DNA barcodes (658 bp) for all *A. crataegi* and one of the four Iranian *A. leucodice* samples, and sequences of various lengths were obtained for the remaining three samples of the Iranian *A. leucodice* (see Supplementary Dataset File).

Phylogenetic analyses in the subtribe Aporiina. The ML tree confirmed the monophyly of *A. crataegi* clade (Bootstrap support, BS = 100) and the close relatedness to a clade (BS = 97) consisting of *Aporia* + *Mesapia* (Fig. 1;^{13,18}). Our findings do not corroborate with the morphological groups proposed by Della Bruna et al.^{9,19} based on shape of the uncus in male genitalia. For example, *A. oberthuri* and *A. goutellei* are closely related in our tree (Fig. 1), while the former has a bifid and the latter a pointed apex (see¹⁸). Our analysis further highlighted

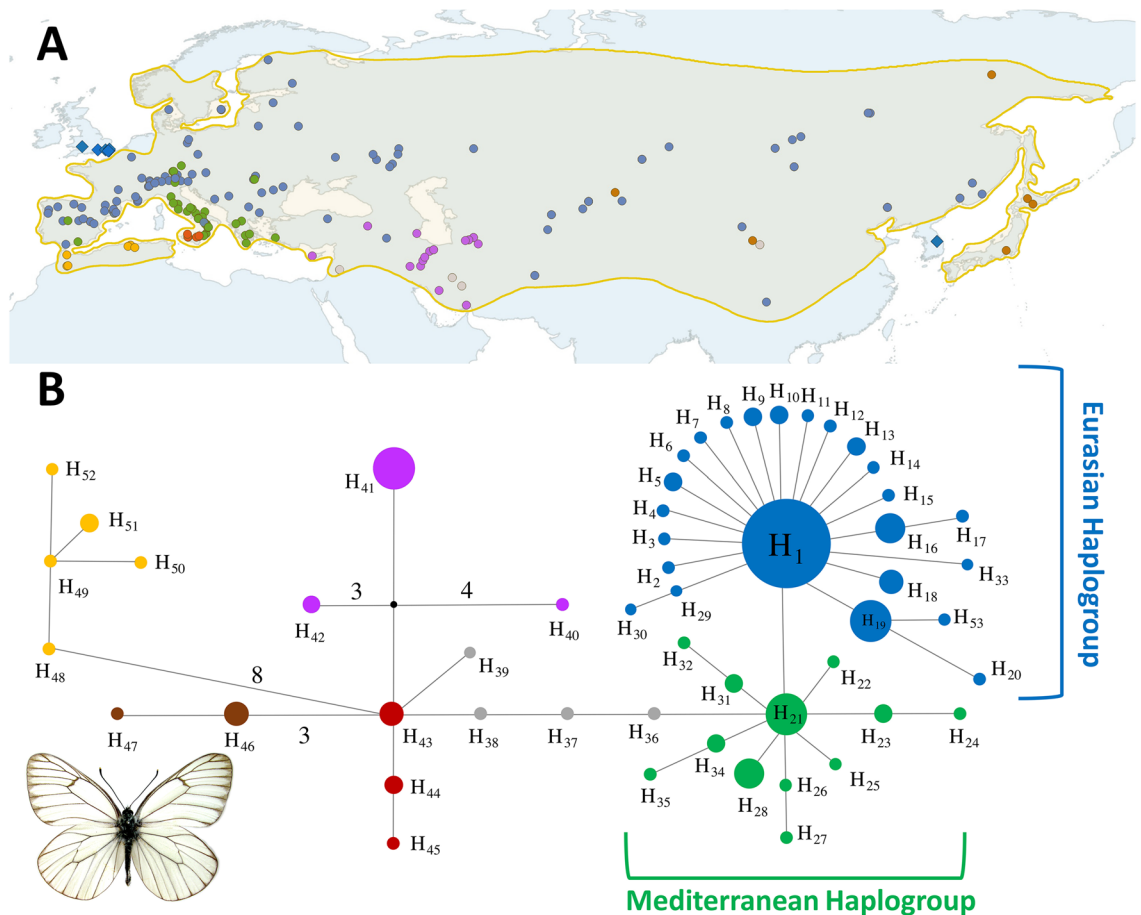


Figure 2. (A) Sampling localities and approximate geographic distribution areas of the *Aporia crataegi* redrawn after^{9,23,25}. Sampling localities for the populations included in this study are showed with different coloured circles, while the extinct populations are indicated with squares. The map was prepared using Quantum GIS 2.8.2 (<https://qgis.org/downloads/>) based on a map from Natural Earth (www.naturalearthdata.com). (B) Median-Joining Network of *A. crataegi* COI sequences. The size of circles is proportional to haplotype frequency and numbers of mutations between haplotypes are shown at the connections, except for single or double substitutions. In both figures main haplogroups are highlighted and shown in different colours. *Aporia crataegi* image is of a specimen from Uzbekistan (UZTAS: Supplementary Dataset File; image courtesy of Josef de Freina).

Tests of demographic equilibrium and estimation of divergence times. The F_s and R_2 statistics, calculated for the Eurasian haplogroup ($N = 114$) and the Mediterranean haplogroup ($N = 35$) sequence sets of *A. crataegi* (Fig. 2B), rejected the null hypothesis of constant population size for these two phylogeographic units (Table 1). Mismatch distribution of these groups was examined according to sudden expansion model (Table 1), and goodness of fit tests did not show significant deviations from expected distributions, so that parameter $\tau = 2\mu t$ could be used to estimate the time (t) elapsed from population expansion. Estimated values of τ and their 5% and 95% confidence limits are shown in Table 1. Based on Brower's⁴⁸ proposed mutation rates in Lepidoptera ($\mu = 0.01$ substitutions/site/lineage/Ma), demographic expansion of the Eurasian and Mediterranean haplogroup could be traced back to about 67 thousand years ago (kya) (49–99 kya) and 116 kya (57–191 kya) respectively.

Discussion

The taxonomic status of the genus *Aporia*. The results of our study based on mitochondrial DNA barcode data alone provide strong support for previous findings based on combined data^{13,18} on the systematic relationships within the genus *Aporia* (Fig. 1). The division of *Aporia* into three subgenera (*Aporia*, *Metaporis* and *Mesapia*) is not supported, since both *Aporia* (*Aporia*) and *Aporia* (*Metaporis*) (sensu¹⁹) are found to be paraphyletic, with *Mesapia* nested within *Aporia* sensu lato. The diagnostic morphological characters for these taxa include smaller wingspan, more rounded wings, and hairy genitalia for *Mesapia*⁴⁹, and minor differences in wing pattern elements for *Metaporis*⁹, none of which, in our opinion, merit a separate subgeneric (sensu^{14,15}) or generic (sensu^{16,17}) status for these taxa. *Metaporis* has already been synonymized with *Aporia*⁹, and our results highlight a need to also synonymize *Mesapia* under *Aporia* (syn. nov). Within *A. leucodice*, the smaller, darker Iranian populations (ssp. *illumina* Grum-Grishimailo, 1890) appeared considerably distinct from the larger, lighter Kazakhstan and Kirghizstan populations (ssp. *leucodice* Eversmann, 1843). The average COI barcode

distance between these two clusters ($2.67 \pm 0.64\%$) falls well within the average distance between most *Aporia* species. Despite this, we refrain from elevating *illumina* as a separate species until supporting ecological or morphological evidence become available. As for *A. joubini*, based on similarities in wing pattern, Della Bruna et al.⁹ suspected that this species may represent an aberrant form of *A. harrietae paracraea*, a closely related taxon that flies in the type locality of *A. joubini*. We agree with this assessment. If barcode sequences of *A. harrietae* become available and prove to be very close or identical to *A. joubini*, the latter should be downgraded to a junior synonym of *A. harrietae*.

Subspecific taxonomy and evolution of *Aporia crataegi*. Within *A. crataegi*, from 23 to 25 subspecies have been described based on fragmented distribution and few morphological characters^{9,10}, but our mtDNA analyses reveal a total of six distinct lineages (Fig. 2). Our data highlighted several populations, described as different subspecies, sharing similar or identical mtDNA sequences, suggesting a very recent evolutionary origin. The most evident example occurs in Eurasia, where 66 samples from 53 localities, described as 12 subspecies according to Della Bruna et al.⁹, share an identical haplotype (H1: Fig. 2B) consistent with recent demographic expansion during the Late Pleistocene glacial period (about 67 kya: Table 1) for the Eurasian haplogroup (Fig. 2B). Populations from three major countries in the Mediterranean Basin (Italy, Greece, and Spain: Fig. 2) also showed very weak differentiation, despite three subspecies (ssp. *rotunda*, ssp. *karavaievi*, and ssp. *rutae* respectively) recognized by Della Bruna et al.⁹, confirming a pervasive phylogeographic pattern in European phylogeography^{50,51}. The trace of demographic expansion model indeed dates a rapid expansion of the Mediterranean haplogroup to about 116 kya during the Late Pleistocene. The sea-level oscillations during this period, which created land bridges between the three main peninsulas⁵² and permitted faunal exchange, can explain the fragmented distribution of this haplogroup (Fig. 2). This scenario has been suggested as a hypothetical distribution pattern during the last glacial period of many organisms including butterflies (e.g. *Maniola jurtina*⁵⁰). Afterwards, in the post glacial period, gene flow among these populations became interrupted, and they experienced a second expansion towards the continent. It is worth noting that although one mutation separates the two main haplogroups, the present postglacial distribution of the Mediterranean haplogroup in *A. crataegi* (Fig. 2A) appears to be an example of the “grasshopper paradigm”⁵⁰, frequently repeated in many animal and plant species^{53,54} including butterflies (e.g. *Zerynthia polyxena*⁵⁵). The grasshopper paradigm is expected for species showing difficulties in crossing the mountain barriers represented by the Alps and Pyrenees^{50,55}. Indeed, in *A. crataegi*, the populations of the Mediterranean haplogroup from Italy and Iberia remained trapped by the Alps and Pyrenees. Conversely, the Greek populations did not expand to Central Europe despite the absence of conspicuous mountain chains that would otherwise impede such colonization. Although the validity of using mtDNA as a marker in molecular ecology has been questioned in the last decade^{56,57}, a recent study has provided evidence that mtDNA spatial differentiation is correlated with species traits known to affect the dispersal and colonization capabilities of butterflies⁵¹. Nevertheless, considering that the mtDNA variation observed in this study is not representative of the whole genomic variation, we cannot exclude a recent expansion of the Greek populations in Central Europe. Nuclear genetic analyses are therefore needed to better investigate the genetic variability, particularly in the contact zones between the main haplogroups. Furthermore, additional sampling might reveal a different genetic pattern in *A. crataegi*.

The extinct British and Korean populations appear to be genetically most closely related to the other specimens from Eurasia (Supplementary File 1). Although this signals possible synonymy of these populations, we refrain from making any changes to their current taxonomy, as any proposal to synonymize a large number of ‘subspecies’ of *A. crataegi* with identical mtDNA haplotypes will require a more in-depth analyses of morphological differences and examination of type materials.

Our samples from five Sicilian populations showed three unique haplotypes, highlighting their significance for biodiversity in Sicily. In Iran, 15 samples from 12 localities encompassing four subspecies (sensu⁹) share an identical haplotype (H41: Fig. 2B). Our analysis strongly supports the distinctiveness of the Iranian populations with the exception of two divergent sequences from Central Iran (Yazd, H36 and Kerman, H37: Supplementary Dataset File). In addition, our results support the distinctiveness of the single sample from Syria (ssp. *augustior*, sensu⁹). As stated above, further sampling and nuclear genetic analyses particularly from Iran, Israel, Syria, Jordan, and Iraq will likely reveal additional haplotypes. Finally, our data supports the genetic peculiarity and evolutionary value of the populations from Japan and north-western Africa, previously recognized as two well-differentiated subspecies (ssp. *adherbal* and ssp. *mauritanica* respectively). Although our phylogenetic analysis doesn't fully resolve the relationships between the North African and Sicilian populations (ssp. *mauritanica* and ssp. *augusta* respectively), the large differentiation between these populations of *A. crataegi* suggests a prolonged period of lack of genetic exchange between them, likely as a result of an early split in the range of the ancestral stock. Despite the small geographical distance between North Africa and Sicily, eight mutations divide these two subspecies (Fig. 2B). The last land bridge between Europe and Africa is known to have occurred at the end of Miocene (7–5.3 million years ago (mya))⁵⁸, when a temporary closure of the water corridors between Africa and the Iberian Peninsula permitted biotic exchange between the two continents⁵⁹. During this period, known as the ‘Messinian Salinity Crisis’, Spain was connected to Morocco, and Tunisia to Italy via Sicily^{60,61}. Re-opening of the Gibraltar at the beginning of the Pliocene (5 mya) restored the barrier. This short period of connection between the two continents has been suggested as a plausible explanation for vicariance between the North African, Iberian, Sicilian and Italian lineages of many organisms, including butterflies (e.g. *Melanargia*⁶²; *Pseudophilotes*⁶³). However, considering Brower's⁴⁸ proposed mutation rates for mtDNA in Lepidoptera ($\mu = 0.01$ substitutions/site/lineage/Ma), demographic expansion of major *Aporia crataegi* haplogroups appear to have a much more recent (Late Pleistocene) origin and are possibly influenced by glaciations or other deterministic and stochastic events, as previously noted in other Lepidoptera (e.g. *Polyommatus*⁶⁴; *Aricia*⁶⁵).

Conclusions

Natural history collections represent an invaluable source of information for species whose ranges are characterized by geomorphological or geopolitical impediments, but principally for storing specimens of extinct taxa when fresh biological material is no longer available, as the case of *A. joubini* and British populations of *A. crataegi*. These repositories allow us to fill the gap in genetic information and reconstruct the evolutionary history of species. For now, NGS-based analysis of COI barcodes has made efficient genomic studies of museum specimens possible, allowing the first look at the genetic traits of extinct species. Rapid advances in modern DNA technology will eventually allow complete reconstruction of genomes, uncovering the lost evolutionary history of these extinct populations and ushering us into a new age of Biological and Conservation Sciences.

Data availability

The datasets generated during the current study are available in the BOLD system repository: Dataset name: DS—APORIA; <https://doi.org/10.5883/DS-APORIA>.

Received: 11 February 2020; Accepted: 7 July 2020

Published online: 18 August 2020

References

- Hughes, A. R. & Stachowicz, J. J. Genetic diversity enhances the resistance of a seagrass ecosystem to disturbance. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8998–9002 (2004).
- Frankham, R. Genetics and extinction. *Biol. Conserv.* **126**, 131–140 (2005).
- Koizumi, I., Usio, N., Kawai, T., Azuma, N. & Masuda, R. Loss of genetic diversity means loss of geological information: The endangered Japanese crayfish exhibits remarkable historical footprints. *PLoS ONE* **7**(3), e33986. <https://doi.org/10.1371/journal.pone.0033986> (2012).
- Nazari, V., Schmidt, B. C., Prosser, S. & Hebert, P. D. N. Century-old DNA barcodes reveal phylogenetic placement of the extinct Jamaican Sunset Moth, *Urania sloanus* Cramer (Lepidoptera: Uraniidae). *PLoS ONE* **11**(10), e0164405. <https://doi.org/10.1371/journal.pone.0164405> (2016).
- Nazari, V., Tarmann, G. & Efetov, K. A. Phylogenetic position of the ‘extinct’ Fijian coconut moth, *Levuana iridescens* (Lepidoptera: Zygaenidae). *PLoS ONE* **14**(12), e0225590. <https://doi.org/10.1371/journal.pone.0225590> (2019).
- Kim, Y. S. *Illustrated Book of Korean Butterflies in Color* (Kyo-Hak Publishing, Seoul, 2002).
- Park, J. S., Cho, Y., Kim, M. J., Nam, S. H. & Kim, I. Description of complete mitochondrial genome of the black-veined white, *Aporia crataegi* (Lepidoptera: Papilionoidea), and comparison to papilionoid species. *J. Asia-Pac. Entomol.* **15**, 331–341 (2012).
- Park, H. C. *et al.* DNA barcode analysis for conservation of an endangered species, *Aporia crataegi* (Lepidoptera, Pieridae) in Korea. *J. Seric. Entomol. Sci.* **51**, 201–206 (2013).
- Della Bruna, C., Gallo, E. & Sbordoni, V. *Pieridae, part I (2nd edition). Guide to the Butterflies of the Palearctic Region* (ed. Bozano, G. C.) (Omnes Artes, 2013).
- Savelle, M. *Lepidoptera and some other life forms* <https://www.nic.funet.fi/pub/sci/bio/life/insecta/lepidoptera/> (2018). Accessed 10 November 2019.
- Braby, M. F., Vila, R. & Pierce, N. E. Molecular phylogeny and systematics of the Pieridae (Lepidoptera: Papilionoidea): Higher classification and biogeography. *Zool. J. Linn. Soc.* **147**, 239–275 (2006).
- Wahlberg, N., Rota, J., Braby, M. F., Pierce, N. E. & Wheat, C. W. Revised systematics and higher classification of pierid butterflies (Lepidoptera: Pieridae) based on molecular data. *Zool. Scr.* **43**, 641–650 (2014).
- Ding, C. & Zhang, Y. Phylogenetic relationships of Pieridae (Lepidoptera: Papilionoidea) in China based on seven gene fragments. *Entomol. Sci.* **20**, 15–23 (2017).
- Klots, A. B. A generic classification of the Pieridae (Lepidoptera) together with a study of the male genitalia. *Entomol. Am.* **12**, 13–242 (1933).
- Braby, M. F. Provisional checklist of genera of the Pieridae (Lepidoptera: Papilionidae). *Zootaxa* **832**, 1–16 (2005).
- Chou, I. *Monograph of Chinese Butterflies, First Volume* (Henan Scientific and Technological Publishing House, Zhengzhou, 1999).
- Wu, C. *Fauna Sinica, Insecta vol. 52, Lepidoptera, Pieridae* (Science Press, Beijing, 2010).
- Ding, C. & Zhang, Y. Phylogenetic relationships of the genera *Aporia* and *Mesapia* (Lepidoptera: Pieridae) based on COI and EF1 α gene sequences. *Acta Entomol. Sin.* **59**(8), 880–887 (2016).
- Della Bruna, C., Gallo, E., Sbordoni, V. & Bozano, G. C. Addenda to the genus *Aporia* Hübner, [1819] (Lepidoptera: Pieridae). *Nachrichten des Entomologischen Vereins Apollo N. F.* **29**(4), 205–209 (2009).
- Deodati, T. Filogenesi molecolare, barcoding e stima dei tempi evolutivi in tre gruppi di farfalle diurne (Insecta: Lepidoptera) a distribuzione Sino-Himalayana (PhD thesis) (University of Rome Tor Vergata, 2010).
- Allan, P. B. M. *Moths and Memories* 1st edn. (Watkins & Doncaster, Pudleston, 1948).
- Jugovic, J., Črne, M. & Lužnik, M. Movement, demography and behaviour of a highly mobile species: A case study of the black-veined white, *Aporia crataegi* (Lepidoptera: Pieridae). *Eur. J. Entomol.* **114**, 113–122 (2017).
- Kim, T. G., Han, Y. G., Kwon, O. & Cho, Y. Changes in *Aporia crataegi*'s potential habitats in accordance with climate changes in the northeast Asia. *J. Ecol. Environ.* **38**(1), 15–23 (2015).
- van Swaay, C. *et al.* *Aporia crataegi*. *The IUCN Red List of Threatened Species*; <https://www.iucnredlist.org> (2014). Accessed 10 November 2019.
- van Swaay, C. *et al.* *Aporia crataegi*. *The IUCN Red List of Threatened Species*; <https://www.iucnredlist.org> (2010). Accessed 10 November 2019.
- Ivanova, N. V., deWaard, J. R. & Hebert, P. D. N. An inexpensive, automation friendly protocol for recovering high-quality DNA. *Mol. Ecol. Notes* **6**, 998–1002 (2006).
- Prosser, S. W. J., de Waard, J. R., Miller, S. E. & Hebert, P. D. N. DNA barcodes from century-old type specimens using next-generation sequencing. *Mol. Ecol. Resour.* **16**(2), 487–497 (2016).
- Hajibabaei, M. *et al.* Critical factors for assembling a high volume of DNA barcodes. *Philos. Trans. R. Soc. Lond. B* **360**, 1959–1967 (2005).
- deWaard, J. R., Ivanova, N. V., Hajibabaei, M. & Hebert, P. D. N. Assembling DNA barcodes: Analytical protocols. *Method Mol. Biol. Environ. Genet.* **410**, 275–293 (2008).
- Librado, P. & Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
- Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**(6), 1547–1549 (2018).

32. Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
33. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
34. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
35. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**(21), 2688–2690 (2006).
36. Boc, A., Diallo, A. B. & Makarenkov, V. T-REX: A web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res.* **40**(W1), W573–W579 (2012).
37. Rambaut, A. *Figtree v1.4*. <https://tree.bio.ed.ac.uk/software/figtree> (2012). Accessed 23 November 2019.
38. Kimura, M. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet. Res.* **11**, 247–269 (1968).
39. Ramos-Onsins, S. & Rozas, J. Statistical properties of new neutrality tests against population growth. *Mol. Biol. Evol.* **19**, 2092–2100 (2002).
40. Fu, Y. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925 (1997).
41. Ramirez-Soriano, A., Ramos-Onsins, S. E., Rozas, J., Calafell, F. & Navarro, A. Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics* **179**, 555–567 (2008).
42. Holsinger K. E., Tajima's D, Fu's F S, Fay and Wu's H, and Zeng et al.'s E. in *Lecture Notes in Population Genetics* (ed. Holsinger, K. E. 2001–2019) 291–296 (University of Connecticut, 2017).
43. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
44. Hudson, R. R. Gene genealogies and the coalescent process. In *Oxford Surveys in Evolutionary Biology* (eds Futuyma, D. J. & Antonovics, J. D.) 1–44 (Oxford University Press, Oxford, 1990).
45. Schneider, S. & Excoffier, L. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: Application to human mitochondrial DNA. *Genetics* **152**, 1079–1089 (1999).
46. Ray, N., Currat, M. & Excoffier, L. Intra-deme molecular diversity in spatially expanding populations. *Mol. Biol. Evol.* **20**, 76–86 (2003).
47. Excoffier, L. Patterns of DNA sequence diversity and genetic structure after a range expansion: Lessons from the infinite-island model. *Mol. Ecol.* **13**, 853–864 (2004).
48. Brower, A. V. Z. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato*, inferred from patterns of mitochondrial DNA evolution. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6491–6495 (1994).
49. Tadokoro, T., Koide, Y. & Hori, K. Description of a new subspecies of *Mesapia peloria* from central Nepal, and taxonomic notes for other subspecies (Lepidoptera, Pieridae). *Lepidoptera Sci.* **65**(2), 51–59 (2014).
50. Schmitt, T. Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Front. Zool.* **4**(1), 11 (2007).
51. Dapporto, L. *et al.* Integrating three comprehensive data sets shows that mitochondrial DNA variation is linked to species traits and paleogeographic events in European butterflies. *Mol. Ecol. Resour.* **00**, 1–14 (2019).
52. Ehlers, J., Astakhov, V., Gibbard, P. L., Mangerud, J. & Svendsen, J. I. Late pleistocene glaciations in Europe. In *Encyclopedia of Quaternary Science* (ed. Elias, S. A.) 1085–1095 (Elsevier, Hoboken, 2006).
53. King, R. A. & Ferris, C. Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Mol. Ecol.* **7**, 1151–1161 (1998).
54. Wallis, G. P. & Arntzen, J. W. Mitochondrial-DNA variation in the crested newt superspecies: Limited cytoplasmic gene flow among species. *Evolution* **43**, 88–104 (1989).
55. Dapporto, L. Speciation in Mediterranean refugia and post-glacial expansion of *Zerynthia polyxena* (Lepidoptera, Papilionidae). *J. Zool. Syst. Evol. Res.* **48**(3), 229–237 (2010).
56. Galtier, N., Nabholz, B., Glémin, S. & Hurst, G. D. D. Mitochondrial DNA as a marker of molecular diversity: A reappraisal. *Mol. Ecol.* **18**(22), 4541–4550 (2009).
57. Stein, E. D., Martinez, M. C., Stiles, S., Miller, P. E. & Zakharov, E. V. Is DNA barcoding actually cheaper and faster than traditional morphological methods: Results from a survey of freshwater bioassessment efforts in the United States?. *PLoS ONE* **9**(4), e95525. <https://doi.org/10.1371/journal.pone.0095525> (2014).
58. Sanmartin, I. Dispersal vs. vicariance in the Mediterranean: historical biogeography of the Palearctic Pachydeminae (Coleoptera, Scarabaeoidea). *J. Biogeogr.* **30**, 1883–1897 (2003).
59. Krijgsman, W. The Mediterranean: Mare nostrum of earth sciences. *Earth Planet. Sci. Lett.* **205**, 1–12 (2002).
60. Steininger, F. F. & Rogl, F. Paleogeography and palinspastic reconstruction of the Neogene of the Mediterranean and Paratethys. In *The Geological Evolution of the Eastern Mediterranean* (eds Dixon, J. E. & Robertson, A. H. F.) 659–668 (Geological Society Special Publication. Blackwell Scientific Publications, Hoboken, 1996).
61. De Jong, H. In search of historical biogeographic patterns in the western Mediterranean terrestrial fauna. *Biol. J. Linn. Soc.* **65**, 99–164 (1998).
62. Nazari, V., Ten Hagen, W. & Bozano, G. C. Molecular systematics and phylogeny of the 'Marbled Whites' (Lepidoptera: Nymphalidae, Satyrinae, *Melanargia* Meigen). *Syst. Entomol.* **35**, 132–147 (2010).
63. Todisco, V. *et al.* Molecular phylogeny of the Palearctic butterfly genus *Pseudophilotes* (Lepidoptera: Lycaenidae) with focus on the Sardinian endemic *P. barbaggiae*. *BMC Zool.* **3**, 4 (2018).
64. Dinca, V., Dapporto, L. & Vila, R. A combined genetic-morphometric analysis unravels the complex biogeographical history of *Polyommatus icarus* and *Polyommatus celina* common blue butterflies. *Mol. Ecol.* **20**, 3921–3935 (2011).
65. Sañudo-Restrepo, C. P., Dinca, V., Talavera, G. & Vila, R. Biogeography and systematics of *Aricia* butterflies (Lepidoptera, Lycaenidae). *Mol. Phylogenet. Evol.* **66**, 369–379 (2013).

Acknowledgements

We are extremely grateful to Josef de Freina (Germany), Alireza Naderi (Iran), J.H. Skevington (Canada), J. Agius (Malta), Alexei G. Belik (Russia), Alex Samus (Russia), Marko Mutanen (University of Oulu) and Geoff Martin (NHM of London) for providing specimens. We also thank Alberto Zilli (NHM of London) for helpful discussions and the Centre for Biodiversity Genomics (CBG) for technical support.

Author contributions

V.T. and V.N. designed the study. V.T., V.N. and R.V. collected samples. S.P. performed NGS analysis and edited the sequences. V.T. analysed the data. V.T. and V.N. wrote the manuscript. All authors edited the manuscript and approved its final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-70957-4>.

Correspondence and requests for materials should be addressed to V.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020