EUROPEAN JOURNAL OF ENTOMOLOGY
ISSN (online): 1802-8829
http://www.eje.cz
Eur. J. Entomol. **119**: 379–387, 2022
doi: 10.14411/eje.2022.039
ORIGINAL ARTICLE

Patterns of DNA barcode diversity in butterfly species (Lepidoptera) introduced to the Nearctic

JACOPO D'ERCOLE[1,2], LEONARDO DAPPORTO[3], B. CHRISTIAN SCHMIDT[4], VLAD DINCĂ[5], GERARD TALAVERA[6], ROGER VILA[7] and PAUL D.N. HEBERT[1,2]

[1] Centre for Biodiversity Genomics, Guelph, Canada; e-mails: jdercole@uoguelph.ca, phebert@uoguelph.ca

[2] Department of Integrative Biology, University of Guelph, Guelph, Canada

[3] ZEN lab, Department of Biology, University of Florence, Sesto Fiorentino, Italy; e-mail: leonardo.dapporto@unifi.it

[4] Canadian National Collection of Insects, Arachnids and Nematodes, Agriculture and Agri-Food Canada, Ottawa, Canada; e-mail: christian.schmidt@agr.gc.ca

[5] Ecology and Genetics Research Unit, University of Oulu, Oulu, Finland; e-mail: vlad.e.dinca@gmail.com

[6] Institut Botànic de Barcelona (IBB), CSIC-Ajuntament de Barcelona, Barcelona, Spain; e-mail: gerard.talavera@csic.es

[7] Institut de Biologia Evolutiva (CSIC-UPF), Barcelona, Spain; e-mail: roger.vila@csic.es

Key words. Introduced species, DNA barcoding, population genetics, North America

Abstract. One of the main consequences of globalization is the intensification of biological introductions. Because of their negative impact on environments, the early detection and monitoring of introduced species through molecular approaches is gaining increased uptake. This study assembles 2,278 DNA barcode records to examine contemporary patterns of sequence variation in mitochondrial cytochrome *c* oxidase I (COI) in five butterfly species introduced to the Nearctic, with a focus on *Pieris rapae* Linnaeus (Lepidoptera: Pieridae) and *Thymelicus lineola* Ochsenheimer (Lepidoptera: Hesperiidae). Parameters of genetic diversity were low (i.e., $h < 0.606$, $\pi < 0.0039$) for Nearctic populations of all analyzed species. Those of *P. rapae* and *T. lineola* showed marked genetic differentiation from their source populations in the Palearctic. Haplotype distributions in their Nearctic populations exposed a starburst pattern with a few common haplotypes known from Palearctic, and infrequent haplotypes diverging from them at only one or two nucleotide sites. Some uncommon haplotypes were only found in the Nearctic suggesting they originated after invasion, while others also occur in the Palearctic. This study provides an example of genetic paradox of invasion, where species often rapidly expand their distribution and become dominant in the new habitat despite their depleted levels of sequence variation.

INTRODUCTION

Species introductions have intensified (Hulme et al., 2009) and are now one of the major contributors to biotic change (Simberloff, 2013). While most researchers oppose introductions because of their potential negative ecological impacts (Simberloff et al., 2012), the eradication of invasive species is often impossible (Davis et al., 2011). However, monitoring programs can aid their early detection and help to document their impacts on native taxa.

As DNA barcoding was proposed for specimen identification (Hebert et al., 2003a), its potential for the efficient detection of quarantine pests was soon recognized (Hebert & Gregory, 2005). Despite this, it has rarely been used to examine patterns of sequence diversity in invasive species (e.g., Valdez-Moreno et al., 2012; Porco et al., 2013). The recent development of DNA barcode reference libraries for the butterfly faunas of North America and Europe (D'Ercole et al., 2021; Dincă et al., 2021) has made it possible to explore patterns of genetic diversity and its relationship to the expansion of introduced butterfly species.

Five butterfly species have been introduced to the Nearctic from the Palearctic. Two, *Aglais io* Linnaeus (Lepidoptera: Nymphalidae) and *A. urticae* Linnaeus (Lepidoptera: Nymphalidae), are closely related and widespread in the Palearctic. *Aglais io* was first recorded in Montreal in 1997 (Handfield, 1999). The only persistent Nearctic population remains there (Nazari, 2018), but it has also been recently reported from south-central Ontario (e.g., www.inaturalist.org/observations/25521945; www.inaturalist.org/observations/41768017) and from both the east coast (e.g., www.inaturalist.org/observations/99733765; https://www.inaturalist.org/observations/95850782; https://www.inaturalist.org/observations/97291558) and west coast (Nazari, 2018). *Aglais urticae* was first reported from Halifax (Scott & Wright, 1972), but most subsequent records derive from New York City (Glassberg, 1992), with a sin-

gle recent report from Florida (e.g., www.inaturalist.org/observations/20129658). The broad geographic and temporal dispersion of these records suggest that both *Aglais* species have been introduced on multiple occasions (Zirlin, 2002). *Polyommatus icarus* Rottemburg (Lepidoptera: Lycaenidae), another common Palearctic butterfly, was first detected near Montreal in 2005 and has expanded its range by about 100 km in 15 years (C. Schmidt, pers. observ.). It has also been reported from Toronto and Quebec City (e.g., www.inaturalist.org/observations/65528682; www.inaturalist.org/observations/26011839), suggesting recurrent introductions. *Pieris rapae* Linnaeus (Lepidoptera: Pieridae) is widely distributed in the Palearctic, where it is considered a pest (Hely et al., 1982). Reflecting its arrival in the latter half of the 19th century, this species has the longest history in the Nearctic where multiple introductions have been reported (Ryan et al., 2019). *Thymelicus lineola* Ochseneimer (Lepidoptera: Hesperiidae), another widespread Palearctic species, is sometimes common enough to be a pest in southern Ontario (Pengelly, 1961; Arthur, 1962). Its first Nearctic specimens were collected in 1910 from London, Ontario (Saunders, 1916), but it is now widespread across the continent (Layberry et al., 1998).

The present study employs DNA barcodes to characterize levels of mitochondrial COI variation in these introduced species. This work aims to document the occurrence of multiple introduction events, identify potential source regions, and to detect novel haplotypes that potentially arose after introduction.

MATERIAL AND METHODS

Sampling

This study assembled DNA barcodes for five butterfly species introduced to the Nearctic. Overall, 2,278 barcode samples (614 from the Nearctic and 1,664 from the Palearctic) were retrieved from BOLD (boldsystems.org) and are assembled in the public dataset DS-INTRSP (dx.doi.org/10.5883/DS-INTRSP). While 60% of these sequences were included in DNA barcode libraries for European (Dincă et al., 2021; Dapporto et al., 2022) and North American (D'Ercole et al., 2021) butterflies, 916 new records were retrieved from BOLD. Among these samples, 101 belonged to *A. io*, 109 to *A. urticae*, 1,095 to *P. rapae*, 611 to *P. icarus*, and 362 to *T. lineola* (Table 1). As no Nearctic barcode records were available for *A. urticae*, comparison with Palearctic populations was not possible.

Genetic analysis

This work examined patterns of sequence diversity in the barcode region of the mitochondrial COI gene. Indices of genetic diversity at the population level, haplotype diversity (h) (Nei, 1987), nucleotide diversity (π) (Tajima, 1983), and number of nucleotide differences (D) (Nei & Li, 1979), were estimated with Arlequin v. 3.5.1.2. (Excoffier & Lischer, 2010). Fasta files were downloaded from BOLD, edited with Geneious ver. 2022.1.1, and converted to arp format with Rstudio ver. 1.3.959 for subsequent Arlequin analysis. Geographic coordinates (WGS84) were retrieved from BOLD and from iodatabase (https://github.com/leondap/iodatabase; Dapporto et al., 2022). Despite the widespread use of h and π, it is difficult to determine the significance of their values because past studies have targeted different taxonomic groups, gene regions, and geographic ranges (Goodall-Copestake et al., 2012). The present work computed average values of haplotype and nucleotide diversity for the butterfly species of North America (D'Ercole et al., 2021) and employed them as a point of reference between low/high values. The Exact Test (Raymond & Rousset, 1995; Goudet et al., 1996) was employed to verify the hypothesis of random distribution of individuals in their native versus introduced range. Pearson's coefficient was employed to assess the correlation between h and species area, as well as h and time since introduction, with the R package ggplot2. The species area for each species was defined by overlaying the geographic locations of observations (https://www.inaturalist.org/observations) on a map of North America. The perimeter delimiting these locations was then drawn to estimate the area of the species distribution.

Statistical approaches, ranging from specialized parametric model-based methods (e.g., Zhang et al., 2010) to more general data-driven non-parametric methods (e.g., Hsieh et al., 2016; Phillips et al., 2020), have recently been proposed to estimate various metrics of genetic diversity. Sampling completeness was assessed using the iNEXT (iNterpolation and EXTrapolation) (Hsieh et al., 2016) R package which generates accumulation curves (Chao, 1984) that can estimate the total haplotype diversity in a species. The asymptotic values of these curves were compared with the observed haplotype diversity to quantify sampling completeness (Dincă et al., 2021). TCS networks (Clement et al., 2000) were generated with PopArt ver. 1.7 (Leigh & Bryant, 2015) to depict the relationships between the Palearctic and Nearctic populations of each species. Nucleotide positions with ambiguities or missing data were masked so network construction only considered sites with full coverage. This reduced consideration to 583 base pairs (bp) for *A. io*, 402 bp for *P. rapae*, 486 bp for *P. icarus*, and 501 bp for *T. lineola*. A BLAST search (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was performed to ascertain if haplotypes detected only in the Nearctic were found outside the study area.

The haplotype maps for each species were constructed using the framework described by Dapporto et al. (2022). Geographic coordinates in decimal degrees were first converted to the World

Table 1. Genetic diversity indices for the five species introduced to North America. N – number of sampled individuals; h – haplotype diversity (Nei, 1987); π – nucleotide diversity (Tajima, 1983); D – average number of nucleotide differences (Nei & Li, 1979); Exact Test (Raymond & Rousset, 1995; Goudet et al., 1996); ND – no available data.

Species	North America				Europe				Populations comparison	
	N	h	π	Shared haplotypes	N	h	π	Shared haplotypes	D	Exact Test (P-values)
Aglais io	3	0	0	1\1	98	0.37	0.0007	1\10	0.23	< 0.001
Aglais urticae	ND	ND	ND	ND	109	0.59	0.0045	ND	ND	ND
Pieris rapae	432	0.51	0.0013	4\9	663	0.99	0.0038	4\47	2.49	< 0.001
Polyommatus icarus	8	0	0	1\1	603	0.93	0.0093	1\53	6.59	< 0.001
Thymelicus lineola	171	0.35	0.0005	4\6	191	0.93	0.0159	4\36	8.95	< 0.001

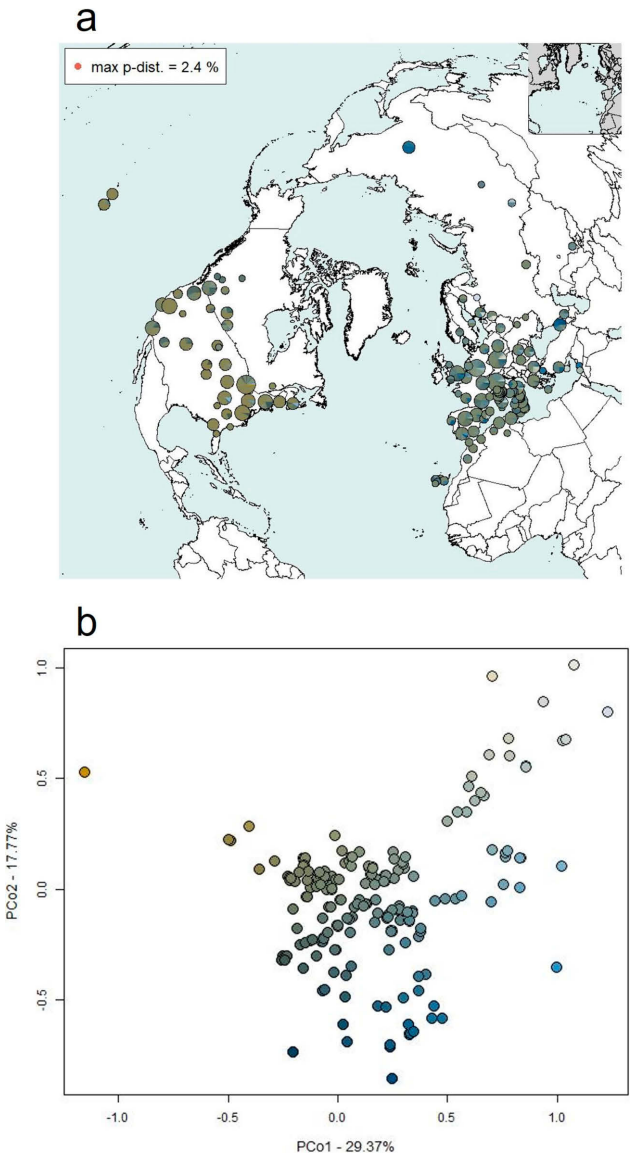

Fig. 1. Haplotype map for 1,095 specimens of *Pieris rapae* (a). Colors match the bidimensional color space of the PCoA projection (b).

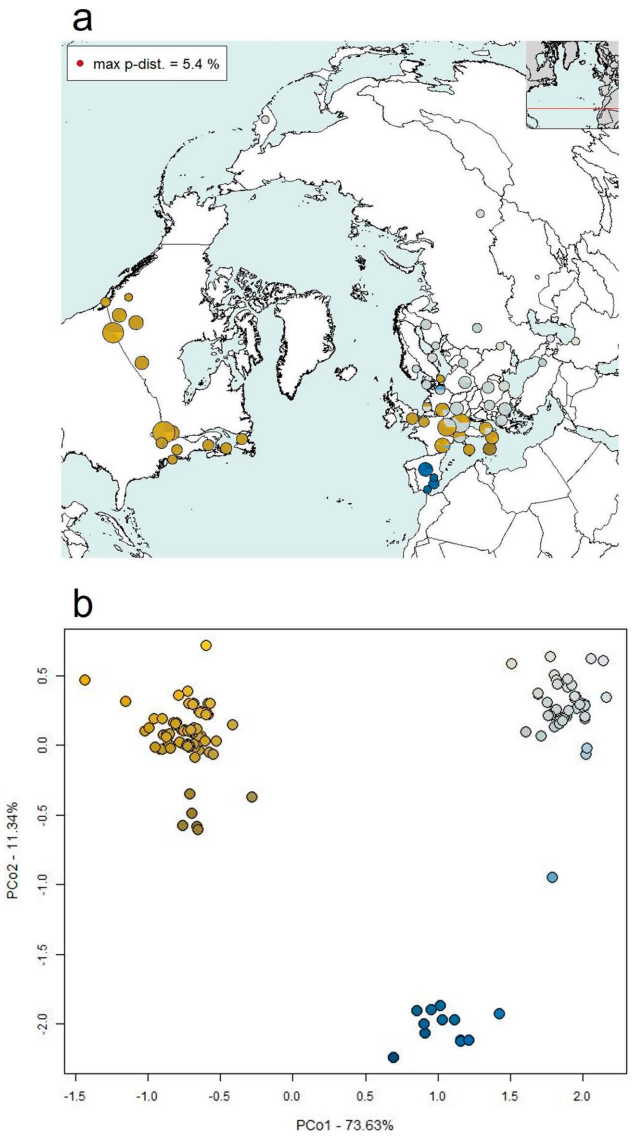

Fig. 2. Haplotype map for 362 specimens of *Thymelicus lineola* (a). Colors match the bidimensional color space of the PCoA projection (b).

Geodetic System (WGS84), and then transformed into Lambert azimuthal equal-area projection to obtain geographic areas of similar size. P-distances among specimens for each species were subjected to a Principal Coordinates Analysis (PCoA). The two-dimensional PCoA was plotted over a square displaying all possible color shades to display genetic distances among specimens. The most common types of color blindness were taken into consideration. Grid cells (500 km by 500 km) were employed to group nearby specimens into a single haplotype pie. The R scripts used to obtain the haplotype maps are available on the iodatabase (https://github.com/leondap/iodatabase; Dapporto et al., 2022).

RESULTS

Sampling

Barcode records were available for Nearctic specimens for four of the introduced species (Figs 1, 2, S1, S2), but sample sizes were so low for two species (*A. io* – 3, *P. icarus* – 8) that sampling completeness was only computed for *P. rapae* and *T. lineola* (Table S1). A total of 16 and 98 unique haplotypes, contributing to 33% and 55% of the total estimated diversity, were obtained respectively for *P. rapae* in the Nearctic and Palearctic. This estimate suggests that 29 haplotypes await sampling in the Nearctic versus 120 in the Palearctic. Estimates of sampling completeness for *T. lineola* showed that 7 and 61 haplotypes were retrieved respectively from the Nearctic and Palearctic, values corresponding to 47% and 30% of their total estimated diversity. This result suggests that 8 haplotypes await discovery in the Nearctic and 144 in the Palearctic (Table S1).

Genetic diversity and haplotype distribution

Values for h and π were consistently low for all Nearctic populations when compared to reference values (h = 0.606, π = 0.0039). This contrasted with the high values for these metrics in Palearctic populations of *P. icarus*, *P. rapae*, and *T. lineola*. By contrast, *A. io* showed low genetic diversity in the Palearctic (Table 1). The Exact Test revealed

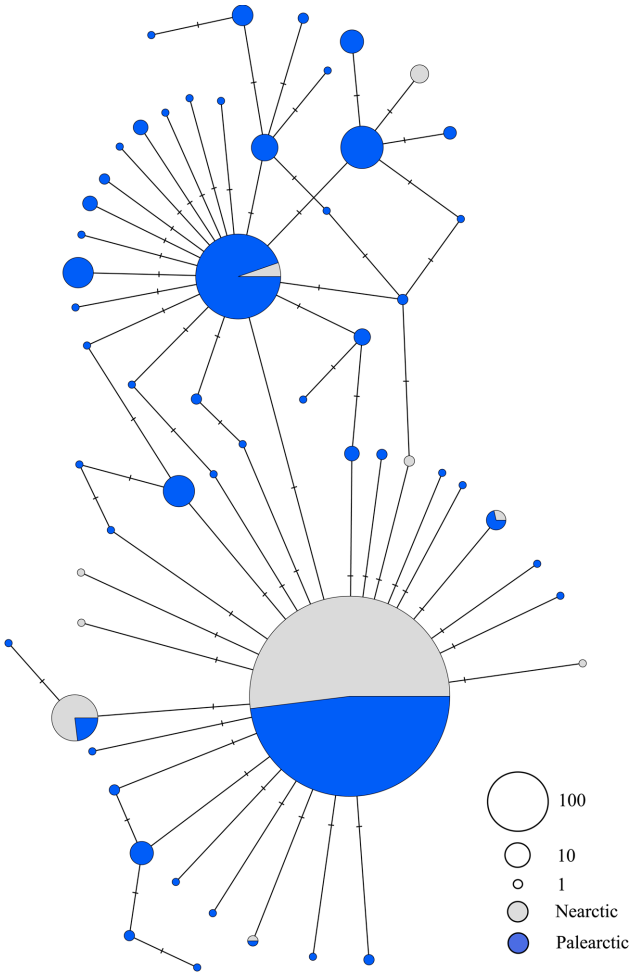

Fig. 3. Haplotype network for 1,095 barcoded specimens of *Pieris rapae*. The number of hypothesized mutational steps is indicated by hatch marks; circle size is proportional to the number of samples.

a non-random distribution of haplotypes in the Nearctic and the Palearctic populations for each of the four species (p-value < 0.001) (Table 1). Pearson's coefficient revealed a moderately strong but non-significant positive correlation between h and geographic range (R = 0.87; p-value = 0.13) (Fig. S3) and a significant correlation between h and time since introduction (P = 1; p-value = 0.0012) (Fig. S4). However, because these correlations are only based on four data points, results need to be interpreted with caution.

The sole Nearctic haplotypes for *A. io* and *P. icarus* were, as expected, also dominant in the Palearctic (Figs S1, S2). Among the ten haplotypes detected in Nearctic *P. rapae*, the dominant one was also abundant in the Palearctic. The other Nearctic haplotypes showed just one or two nucleotide substitutions from those in the Palearctic, but five were only detected in the Nearctic (Fig. 3). Two of the seven Nearctic haplotypes of *T. lineola* were common and known from the Palearctic. The other five haplotypes were infrequent, and three of them were only found only in the Nearctic (Fig. 4). A BLAST search for the haplotypes detected only in the Nearctic confirmed they were exclusive to this continent.

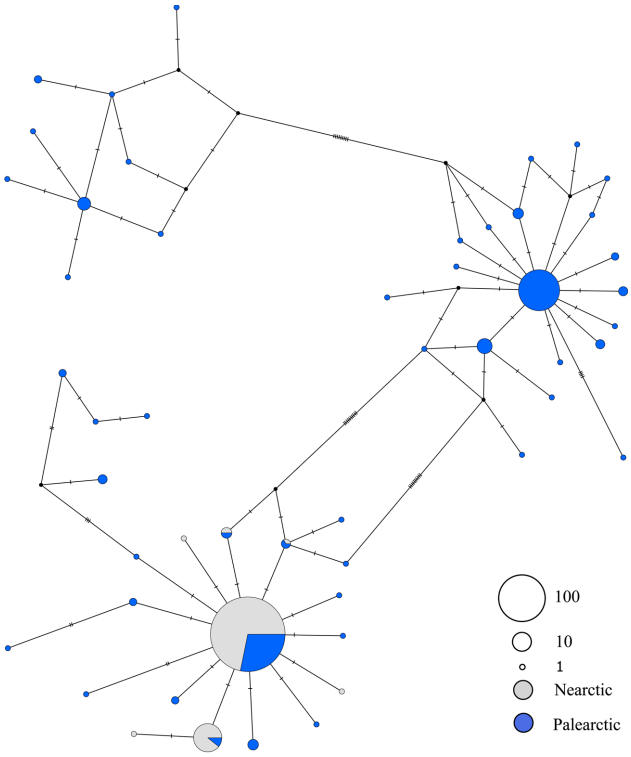

Fig. 4. Haplotype network for 364 barcoded specimens of *Thymelicus lineola*. The number of hypothesized mutational steps is indicated by hatch marks; circle size is proportional to the number of samples.

DISCUSSION

The capacity of DNA barcoding to both assign specimens to a species (Hebert et al., 2003b) and to reveal intraspecific variation (e.g., Hebert et al., 2004; Burns et al., 2007; Dapporto et al., 2022) motivated its present use to compare patterns of genetic diversity in populations of butterfly species introduced to the Nearctic with those in their ancestral Palearctic populations. While a few barcode records can provide an overview of intraspecific diversity, many more specimens need to be analyzed to obtain a detailed understanding of large-scale patterns of intraspecific variation (Zhang et al., 2010; Phillips et al., 2019). Although sequences for 432 Nearctic specimens were available for *P. rapae*, analysis indicated that they only represented 33% of its estimated haplotype diversity. Fewer samples (171) were available for *T. lineola*, but they represented 47% of its estimated haplotype diversity. The limited distribution and coverage for the *A. io* and *P. icarus* prevented their detailed analysis. However, a single haplotype was detected in the Nearctic for both, and these haplotypes were prevalent in the Palearctic as expected for recent introductions. DNA barcodes for *A. urticae* were unavailable in the Nearctic.

Introductions often involve a population bottleneck that results in reduced genetic diversity (Baker & Stebbins, 1965). Although theory suggests such populations are at enhanced risk of extinction (Gilpin & Soulé, 1986; Lacy, 1987; Frankham, 1998), many introduced species thrive — a phenomenon termed the "genetic paradox of invasion"

(Allendorf & Lundquist, 2003). Nearctic populations of *P. rapae* and *T. lineola* follow this pattern as they are common and widely distributed despite their low genetic diversity. Specific ecological features can however outweigh the negative effect of decreased genetic diversity (Estoup et al., 2016). These two species are generalists that feed on diverse plants within a family (Brassicaceae and Poaceae, respectively), tolerate varied climatic regimes (Tolman & Lewington, 2008), and are also adapted to anthropic habitats (Hufbauer et al., 2012). Larvae of *P. rapae* feed on agricultural crops (*Brassica* spp.) as well as many brassicaceous weeds common in anthropic landscapes (Lafranchis, 2007; Tolman & Lewington, 2008). Similarly, *T. lineola* favors timothy grass (*Phleum pratense*) as a larval host plant (Lafranchis, 2007; Tolman & Lewington, 2008). *Polyommatus icarus* has also spread in disturbed landscapes, gaining a foothold via the European host plant *Lotus corniculatus* (Rivest & Kharouba, 2021). Introduced populations can also experience reduced purifying selection because they often host fewer parasitoids and have fewer predators than populations in their native range (Torchin et al., 2003; Perkins et al., 2008). As a result, they can invest more in growth rate and dispersal (Muller-Scharer et al., 2004). It is worth emphasizing that conclusions derived from a single gene marker describe only part of the evolutionary history of an organism. As a result, the reduced diversity observed at mtDNA might not mirror diversity at nuclear loci (Petit-Marty et al., 2021). Not only is mtDNA more exposed to population bottlenecks because of lower number of gene copies, but it is also prone to selective sweeps because of its near absence of recombination (Leffler et al., 2012). It is therefore possible that adaptive potential is maintained, despite the reduced diversity observed at mtDNA.

In agreement with earlier work on *P. rapae* (Scudder, 1886; Andow et al., 1990; Ryan et al., 2019) and *T. lineola* (Burns, 1966), the present results suggest multiple introductions. Because each introduction likely involved just a few closely related individuals from a single Palearctic locality, the presence of multiple widespread haplotypes shared with populations in the native range supports recurrent introductions. Aside from a few common haplotypes, Nearctic populations of *P. rapae* and *T. lineola* possessed infrequent haplotypes one or two mutations away from the dominant lineages. This starburst pattern is the rule in animal lineages (e.g., Chen et al., 2004; Bollongino et al., 2006; Lait & Hebert, 2018) including butterflies (Dincă et al., 2021), but the detection of haplotypes private to the Nearctic is surprising. While their presence in the Palearctic may have been overlooked due to incomplete sampling and introduction from other regions of the world cannot be excluded, they may also reflect variants that arose after introduction. Based on standard rates of molecular evolution of 1.5% per million years (Quek et al., 2004), de novo variants would not be expected given the brief interval since introduction. However, because of the reduced impact of selection and drift at the population level over short time scales, mutation rates for mitochondrial DNA can vary 100-fold from the slow phylogenetic rate to the highest pedigree-based rate (Ho et al., 2005). While this shift might be responsible for the newly arisen substitutions, other factors including increased positive selection (Gillespie, 2001), relaxed purifying selection (Henn, 2009), and gene surfing associated with population expansion (Excoffier & Ray, 2008) can aid the fixation of new variants.

Despite low genetic diversity in their Nearctic populations, positive linear relationships were detected between the time since introduction and geographic range. In agreement with the preceding considerations, it is possible that a combination of multiple introduction events and newly arisen mutations has led to increased diversity through time. Because the geographic area occupied is expected to co-vary with time for populations increasing in size, the observed correlation between genetic diversity and area is not surprising. Moreover, the occupancy of larger areas can, in itself, promote adaptive ecological changes and increase diversity. While this association is well-established for species diversity, the impact of area on genetic diversity is less clear (Dapporto et al., 2019; Lawrence et al., 2020).

Identifying source populations in their native range and admixture of different lineages in the novel environments is critical to determine the capacity to disperse and the evolutionary potential of introduced species (Colautti & Lau, 2015; Dlugosh et al., 2015). Such information is relevant because admixture following multiple introductions from different source regions produces both an increase in genetic diversity and a progressive increase in fitness as a result of dominance (Davenport, 1908) and over-dominance (East, 1908; Shull, 1908). Because this study revealed no clear spatial genetic structure for *P. rapae* in the Palearctic, it was not possible to identify the source region for the Nearctic populations. The same pattern was observed in a study that examined nuclear genomic diversity (Ryan et al., 2019). The evidence for the lack of structure is likely explained by the high dispersal capacity (Dapporto et al., 2019) and human-mediated transport (Jones et al., 1980). By comparison, *Thymelicus lineola* displayed three geographically distinct haplogroups in the Palearctic (Dapporto et al., 2022), and only one was found in the Nearctic. This likely excludes introductions from the Iberian Peninsula, eastern Europe, and Asia. As expected for recent introductions, a single haplotype widespread in the native range was detected in the Nearctic for *A. io* and *P. icarus*. While it is difficult to identify potential source regions for *A. io* due to the lack of clear genetic structure in its native range, *P. icarus* is highly structured in Europe (Dincă et al., 2011; Arif et al., 2021) and the most probable origin for the North American population is anywhere but the Iberian Peninsula. Continuous monitoring of these species will be valuable as the evolution of invading populations is rapid and unpredictable immediately after introduction. Not only a lag time often exists between the initial colonization and the onset of population expansion (Sakai et al., 2001; Bock et al., 2015), but abrupt changes in population size can also occur as a result of varying selective pressure shortly after colonization (Phillips et al., 2010).

This study has compared patterns of genetic diversity in four butterfly species introduced to the Nearctic with those of their Palearctic ancestral populations. The two most recently introduced species, *A. io* and *P. icarus* possess a single haplotype widespread in Europe. While levels of genetic diversity were much lower in Nearctic than Palearctic populations of *P. rapae* and *T. lineola*, both have broad distributions. They also displayed a starburst haplotype network in the Nearctic with a few widespread haplotypes and infrequent ones with one or two mutations. Some rare haplotypes were only found in the Nearctic, suggesting the possibility of their origin after introduction. Supporting conclusions based on their disjunct distributional patterns in the Nearctic, the analysis of haplotype variation for *P. rapae* and *T. lineola* suggest multiple introductions. Analysis of sequence variation in the native range also identifies potential source regions for *T. lineola* and *P. icarus*. Overall, this study demonstrates the effectiveness of DNA barcoding for describing the diversity of introduced species and establish its value to monitoring programs of invasive species.

ACKNOWLEDGEMENTS. We thank D. Szaboova, A. Johnsen, A. Hausmann, J. Reil, M. Seidensticker, J. Kjaerandsen, A. Segerer, V. Lukhtanov, M. Franzen for granting use of their DNA barcode records.

FUNDING. This study was supported, in part, by a NSERC Discovery grant to PDNH and by an award from the Canada First Research Excellence Fund for the support of Food from Thought research program. This research was also supported by the Academy of Finland (Academy Research Fellow, decision no. 328895) to VD, by project PID2020-117739GA-I00/MCIN/AEI/ 10.13039/501100011033 to GT, by project PID2019-107078GB-I00/MCIN/AEI/ 10.13039/501100011033 to RV, and by projects related to the Direttiva Biodiversità del Ministero della Transizione Ecológica to LD.

COMPETING INTERESTS. The authors declare no competing interest.

REFERENCES

Allendorf F.W. & Lundquist L.L. 2003: Introduction: population biology, evolution, and control of invasive species. — *Conserv. Biol*. **17**: 24–30.

Andow D.A., Kareiva P.M., Levin S.A. & Okubo A. 1990: Spread of invading organisms. — *Landsc. Ecol*. **4**: 177–188.

Arif S., Gerth M., Hone-Millard W.G., Nunes M.D., Dapporto L. & Shreeve T.G. 2021: Evidence for multiple colonisations and Wolbachia infections shaping the genetic structure of the widespread butterfly *Polyommatus icarus* in the British Isles. — *Mol. Ecol*. **30**: 5196–5213.

Arthur A.P. 1962: A skipper *Thymelicus lineola* (Ochs.) (Lepidoptera: Hesperiidae) and its parasites in Ontario. — *Can. Entomol*. **94**: 1082–1089.

Baker H. & Stebbins G. 1965: *The Genetics of Colonizing Species*. Academic Press, New York, 458 pp.

Bock D.G., Caseys C., Cousens R.D., Hahn M.A., Heredia S.M., Hübner S., Turner K.G., Whitney K.D. & Rieseberg L.H. 2015: What we still don't know about invasion genetics. — *Mol. Ecol*. **24**: 2277–2297.

Bollongino R., Edwards C.J., Alt K.W., Burger J. & Bradley D.G. 2006: Early history of European domestic cattle as revealed by ancient DNA. — *Biol. Lett*. **2**: 155–159.

Burns J.M. 1966: Expanding distribution and evolutionary potential of *Thymelicus lineola* (Lepidoptera: Hesperiidae), an introduced skipper, with special reference to its appearance in British Columbia. — *Can. Entomol*. **98**: 859–866.

Burns J.M., Janzen D.H., Hajibabaei M., Hallwachs W. & Hebert P.D.N. 2007: DNA barcodes of closely related (but morphologically and ecologically distinct) species of skipper butterflies (Hesperiidae) can differ by only one to three nucleotides. — *J. Lepidopt. Soc*. **61**: 138–153.

Chao A. 1984: Nonparametric estimation of the number of classes in a population. — *Scand. J. Stat.* **11**: 265–270.

Chen C.A., Ablan M.C., McManus J.W., Bell J.D., Tuan V.S., Cabanban A.S. & Shao K.T. 2004: Population structure and genetic variability of six bar wrasse (*Thallasoma hardwicki*) in northern south China Sea revealed by mitochondrial control region sequences. — *Marine Biotechnol*. **6**: 312–326.

Clement M., Posada D. & Crandall K.A. 2000: TCS: a computer program to estimate gene genealogies. — *Mol. Ecol*. **9**: 1657–1659.

Colautti R.I. & Lau J.A. 2015: Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. — *Mol. Ecol*. **24**: 1927–2297.

Cooper A. 1994: DNA from museum specimens. In Herrmann B. & Hummel S. (eds): *Ancient DNA*. Springer, New York, pp. 149–165.

D'Ercole J., Dincă V., Opler P.A., Kondla N.G., Schmidt C., Phillips J.D., Robbins R., Burns J.M., Miller S.E., Grishin N.V. et al. 2021: A DNA barcode library for the butterflies of North America. — *PeerJ*. **9**: e11157, 22 pp.

Dapporto L., Cini A., Vodă R., Dincă V., Wiemers M., Menchetti M., Magini G., Talavera G., Shreeve T., Bonelli S. et al. 2019: Integrating three comprehensive data sets shows that mitochondrial DNA variation is linked to species traits and paleogeographic events in European butterflies. — *Mol. Ecol. Resour*. **19**: 1623–1636.

Dapporto L., Menchetti M., Vodă R., Corbella C., Cuvelier S., Djemadi, I., Gascoigne-Pees M., Hinojosa J.C., Lam N.T., Serracanta M. et al. 2022: The atlas of mitochondrial genetic diversity for Western Palearctic butterflies. — *Global Ecol. Biogeogr*. (prepublished online), 7 pp. DOI: 10.1111/geb.13579.

Davenport C.B. 1908: Degeneration, albinism and inbreeding. *Science* **28**: 454–455.

Davis M.A., Chew M.K., Hobbs R.J., Lugo A.E., Ewel J.J., Vermeij G.J., Brown J.H., Rosenzweig M.L., Gardener M.R., Carroll S.P. et al. 2011: Don't judge species on their origins. — *Nature* **474**: 153–154.

Dincă V., Dapporto L. & Vila R. 2011: A combined genetic-morphometric analysis unravels the complex biogeographical history of *Polyommatus icarus* and *Polyommatus celina* Common Blue butterflies. — *Mol. Ecol*. **20**: 3921–3935.

Dincă V., Dapporto L., Somervuo P., Vodă R., Cuvelier S., Gascoigne-Pees M., Huemer P., Mutanen M., Hebert P.D.N. & Vila R. 2021: High resolution DNA barcode library for European butterflies reveals continental patterns of mitochondrial genetic diversity. — *Commun. Biol*. **4**: 315, 11 pp.

Dlugosch K.M., Anderson S.R., Braasch J., Cang A. & Gillette D. 2015: The devil is in the details: genetic variation in introduced populations and its contributions to invasion. — *Mol. Ecol*. **24**: 2095–2111.

East E.M. 1908: *Inbreeding in corn.* In: *Reports of the Connecticut Agricultural Experiment Station for Years, 1907–1908.*

Connecticut Agricultural Experiment Station, New Haven, pp. 419–429.

Excoffier L. & Ray N. 2008: Surfing during population expansions promotes genetic revolutions and structuration. — *Trends Ecol. Evol*. **23**: 347–351.

Excoffier L. & Lischer H.E.L. 2010: 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.

Frankham R. 1998: Inbreeding and extinction: island populations. — *Conserv. Biol*. **12**: 665–675.

Gillespie J.H. 2001: Is the population size of a species relevant to it evolution? — *Evolution* **55**: 2161–2169.

Gilpin M.E. & Soulé M.E. 1986: Minimum viable populations: processes of extinction. In Soulé M.E. (ed.): *Conservation Biology: The Science of Scarcity and Diversity*. Sinauer, Sunderland, MA, pp. 13–34.

Glassberg J. 1992: *Aglais urticae* (Nymphalidae): a nascent population in North America. — *J. Lepidopt. Soc*. **46**: 302–304.

Goodall-Copestake W.P., Tarling G.A. & Murphy E.J. 2012: On the comparison of population-level estimates of haplotype and nucleotide diversity: a case study using the gene cox1 in animals. — *Heredity* **109**: 50–56.

Goudet J., Raymond M., de Meeüs T. & Rousset F. 1996: Testing differentiation in diploid populations. — *Genetics* **144**: 1933–1940.

Handfield L. 1999: *Les Papillons du Québec.* Broquet, Saint-Constant, 482 pp.

Hebert P.D.N. & Gregory T.R. 2005: The promise of DNA barcoding for taxonomy. — *Syst. Biol*. **54**: 852–859.

Hebert P.D.N., Cywinska A., Ball S.L. & deWaard JR. 2003a: Biological identifications through DNA barcodes. — *Proc. R. Soc. (B)* **270**: 313–321.

Hebert P.D.N., Ratnasingham S. & de Waard J.R. 2003b: Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. — *Proc. R. Soc. (B)* **270**: S96–S99.

Hebert P.D.N., Penton E.H., Burns J.M., Janzen D.H. & Hallwachs W. 2004: Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. — *Proc. Natn. Acad. Sci. USA* **101**: 14812–14817.

Hebert P.D.N., deWaard J.R., Zakharov E.V., Prosser S.W.J., Sones J.E., McKeown J.T.A., Mantle B. & La Salle J. 2013: A DNA 'barcode blitz': rapid digitization and sequencing of a natural history collection. — *PLoS ONE* **8**: e68535, 14 pp.

Hely P.C., Pasfield G. & Gellatley J.G. 1982: *Insect Pests of Fruit and Vegetables in NSW*. International Specialized Book Services, Sydney, New South Wales, 312 pp.

Henn B.M., Gignoux C.R., Feldman M.W. & Mountain J.L. 2009: Characterizing the time dependency of human mitochondrial DNA mutation rate estimates. — *Mol. Biol. Evol*. **26**: 217–230.

Ho S.Y.W., Phillips M.J., Cooper A. & Drummond A.J. 2005: Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. — *Mol. Biol. Evol*. **22**: 1561–1568.

Hsieh T.C., Ma K.H. & Chao A. 2016: iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). — *Meth. Ecol. Evol*. **7**: 1451–1456.

Hufbauer R.A., Facon B., Ravigné V., Turgeon J., Foucaud J., Lee C.E., Rey O. & Estoup A. 2012: Anthropogenically induced adaptation to invade (AIAI): contemporary adaptation to human-altered habitats within the native range can promote invasions. — *Evol. Appl*. **5**: 89–101.

Hulme P.E. 2009: Trade, transport and trouble: managing invasive species pathways in an era of globalization. — *J. Appl. Ecol*. **46**: 10–18.

Ivanova N.V., deWaard J.R. & Hebert P.D.N. 2006: An inexpensive, automation-friendly protocol for recovering high-quality DNA. — *Mol. Ecol. Notes* **6**: 998–1002.

Jones R.E., Gilbert N., Guppy M. & Nealis V. 1980: Long-distance movement of *Pieris rapae*. — *J. Anim. Ecol*. **49**: 629–642.

Lacy R.C. 1987: Loss of genetic diversity from managed populations: interacting effects of drift, mutation, immigration, selection, and population subdivision. — *Conserv. Biol*. **1**: 143–158.

Lafranchis T. 2007: *Papillons d'Europe.* Diatheo, Paris, 380 pp.

Lait L.A. & Hebert P.D.N. 2018: A survey of molecular diversity and population genetic structure in North American clearwing moths (Lepidoptera: Sesiidae) using cytochrome *c* oxidase I. — *PLoS ONE* **13**: e0202281, 16 pp.

Lawrence E.R. & Fraser D.J. 2020: Latitudinal biodiversity gradients at three levels: linking species richness, population richness and genetic diversity. — *Global Ecol. Biogeogr*. **29**: 770–788.

Layberry R., Hall P. & Lafontaine D. 1998: *The Butterflies of Canada*. University of Toronto Press, Toronto, Ontario, 354 pp.

Leffler E.M., Bullaughey K., Matute D.R., Meyer W.K., Ségurel L., Venkat A., Andolfatto P. & Przeworski M. 2012: Revisiting an old riddle: what determines genetic diversity levels within species? — *PLoS Biol*. **10**: e1001388, 9 pp.

Leigh J.W. & Bryant D. 2015: Popart: full-feature software for haplotype network construction. — *Meth. Ecol. Evol*. **6**: 1110–1116.

Muller-Scharer H., Schaffner U. & Steinger T. 2004: Evolution in invasive plants: implications for biological control. — *Trends Ecol. Evol*. **19**: 417–422.

Nazari V., Handfield L. & Handfield D. 2018: The European Peacock butterfly, *Aglais io* (Linnaeus, 1758) in North America (Lepidoptera: Nymphalidae). — *News Lepidopt. Soc*. **60**: 128–129.

Nei M. 1987: *Molecular Evolutionary Genetics*. Columbia University Press, New York, 512 pp.

Nei M. & Li W.H. 1979: Mathematical model for studying genetic variation in terms of restriction endonucleases. — *Proc. Natn. Acad. Sci. USA* **76**: 5269–5273.

Pengelly D.H. 1961: *Thymelicus lineola* (Ochs.) (Lepidoptera: Hesperiidae) a pest of hay and pasture grasses in southern Ontario. — *Proc. Entomol. Soc. Ont*. **91**[1960]: 189–197.

Perkins S.E., Altizer S., Bjornstad O., Burdon J.J., Clay K., Gómez-Aparicio L., Jeschke J.M., Johnson P.T.J., Lafferty K.D., Malmstrom C.M. et al. 2008: Invasion biology and parasitic infections. In Ostfeld R.S., Keesing F. & Eviner V.T. (eds): *Infectious Disease Ecology: The Effects of Ecosystems on Disease and Disease on Ecosystems*. Princeton University Press, New Jersey, pp. 179–204.

Petit-Marty N., Vázquez-Luis M. & Hendriks I.E. 2021: Use of the nucleotide diversity in COI mitochondrial gene as an early diagnostic of conservation status of animal species. — *Conserv. Lett*. **14**: e12756, 7 pp.

Phillips B.L., Brown G.P. & Shine R. 2010: Life-history evolution in range-shifting populations. — *Ecology* **91**: 1617–1627.

Phillips J.D., Gillis D.J. & Hanner R.H. 2019: Incomplete estimates of genetic diversity within species: implications for DNA barcoding. — *Ecol. Evol*. **9**: 2996–3010.

Phillips J.D., French S.H., Hanner R.H. & Gillis D.J. 2020: HACSim: an R package to estimate intraspecific sample sizes for genetic diversity assessment using haplotype accumulation curves. — *PeerJ Computer Sci*. **6**: e243, 37 pp.

Porco D., Decaëns T., Deharveng L., James S.W., Skarżyński D., Erséus C., Butt K.R., Richard B. & Hebert P.D.N. 2013: Biological invasions in soil: DNA barcoding as a monitoring tool in a multiple taxa survey targeting European earthworms and springtails in North America. — *Biol. Invas*. **15**: 899–910.

Quek S.P., Davies S.J., Itino T. & Pierce N.E. 2004: Codiversification in an ant-plant mutualism: stem texture and the evolution of host use in *Crematogaster* (Formicidae: Myrmicinae) inhabitants of *Macaranga* (Euphorbiaceae). — *Evolution* **58**: 554–570.

Ratnasingham S. & Hebert P.D.N. 2007: BOLD: the barcode of life data system. — *Mol. Ecol. Notes* **7**: 355–364.

Raymond M. & Rousset F. 1995: An exact test for population differentiation. — *Evolution* **49**: 1280–1283.

Rivest S.A. & Kharouba H.M. 2021: Anthropogenic disturbance promotes the abundance of a newly introduced butterfly, the European common blue (*Polyommatus icarus*; Lepidoptera: Lycaenidae), in Canada. — *Can. J. Zool*. **99**: 642–652.

RStudio Team 2016: *RStudio: Integrated Development for R*. RStudio Inc., Boston, URL: https://www.rstudio.com.

Ryan S.F., Lombaert E., Espeset A., Vila R., Talavera G., Dincă V., Doellman M.M., Renshaw M.A., Eng M.W., Hornett E.A. et al. 2019: Global invasion history of the agricultural pest butterfly *Pieris rapae* revealed with genomics and citizen science. — *Proc. Natn. Acad. Sci. USA* **116**: 20015–20024.

Saitou N. & Nei M. 1987: The neighbor-joining method: a new method for reconstructing phylogenetic trees. — *Mol. Biol. Evol*. **4**: 406–425.

Sakai A., Allendorf F., Holt J.S., Lodge D.M., Molofsky J., With K.A., Baughman S., Cabin R.J., Cohen J.E., Ellstrand N.C. et al. 2001: The population biology of invasive species. — *Annu. Rev. Ecol. Syst*. **32**: 305–332.

Saunders W.E. 1916: European butterfly found at London, Ont. — *Ottawa Nat*. **30**: 116.

Scudder S.H. 1887: Introduction and spread of *Pieris rapae* in North America, 1860–1885. — *Mem. Boston Soc. Natl. Hist*. **4**: 53–69.

Shull G.H. 1908: The composition of a field of maize. — *J. Heredity* **4**: 296–301.

Simberloff D., Souza L., Nuñez M.A., Barrios-Garcia M.N. & Bunn W. 2012: The natives are restless, but not often and mostly when disturbed. — *Ecology* **93**: 598–607.

Simberloff D., Martin J.L., Genovesi P., Maris V., Wardle D.A., Aronson J., Courchamp F., Galil B., García-Berthou E., Pascal M. et al. 2013: Impacts of biological invasions: what's what and the way forward. — *Trends Ecol. Evol*. **28**: 58–66.

Tajima F. 1983: Evolutionary relationship of DNA sequences in finite populations. — *Genetics* **105**: 437–460.

Tolman T. & Lewington R. 2008: *Collins Butterfly Guide: The Most Complete Guide to the Butterflies of Britain and Europe*. HarperCollins, London, 384 pp.

Torchin M.E., Lafferty K.D., Dobson A.P., McKenzie V.J. & Kuris A.M. 2003: Introduced species and their missing parasites. — *Nature* **421**: 628–630.

Valdez-Moreno M., Quintal-Lizama C., Gómez-Lozano R. & García-Rivas M. 2012: Monitoring an alien invasion: DNA barcoding and the identification of lionfish and their prey on coral reefs of the Mexican Caribbean. — *PLoS ONE* **7**: e36636, 8 pp.

Zhang A.B., He L.J., Crozier R.H., Muster C. & Zhu C.D. 2010: Estimating sample sizes for DNA barcoding. — *Mol. Phylogen. Evol*. **54**: 1035–1039.

Zirlin H. 2002: Strangers in a strange land. — *Am. Butterflies* **10**: 4–11.

Received May 19, 2022; revised and accepted September 13, 2022
Published online September 26, 2022

Table S1. Estimates of sampling completeness. N – number of specimens analyzed; H – number of observed unique haplotypes; R – estimated proportion of haplotype diversity; T – number of total estimated haplotypes.

Species	Locality	N	H	R	T
Pieris rapae	North America	432	16	0.33	45
	Europe	663	98	0.55	178
Thymelicus lineola	North America	171	7	0.47	15
	Europe	191	61	0.30	205

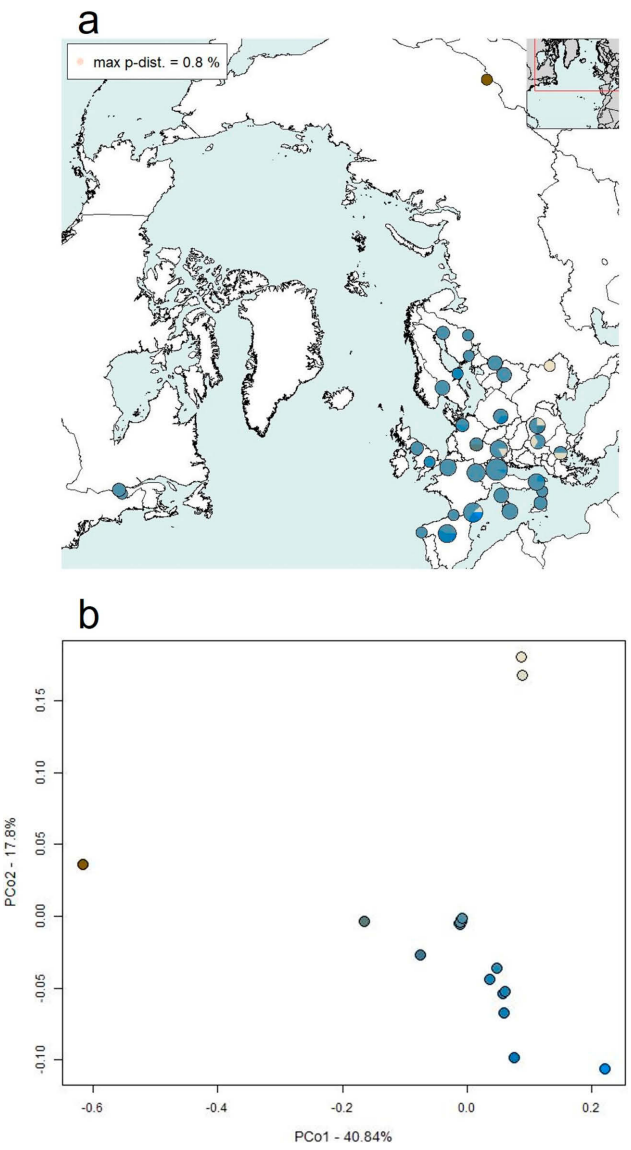

Fig. S1. Haplotype map for 101 specimens of *Aglais io* (a). Colors match the bidimensional color space of the PCoA projection of p-distances (b).

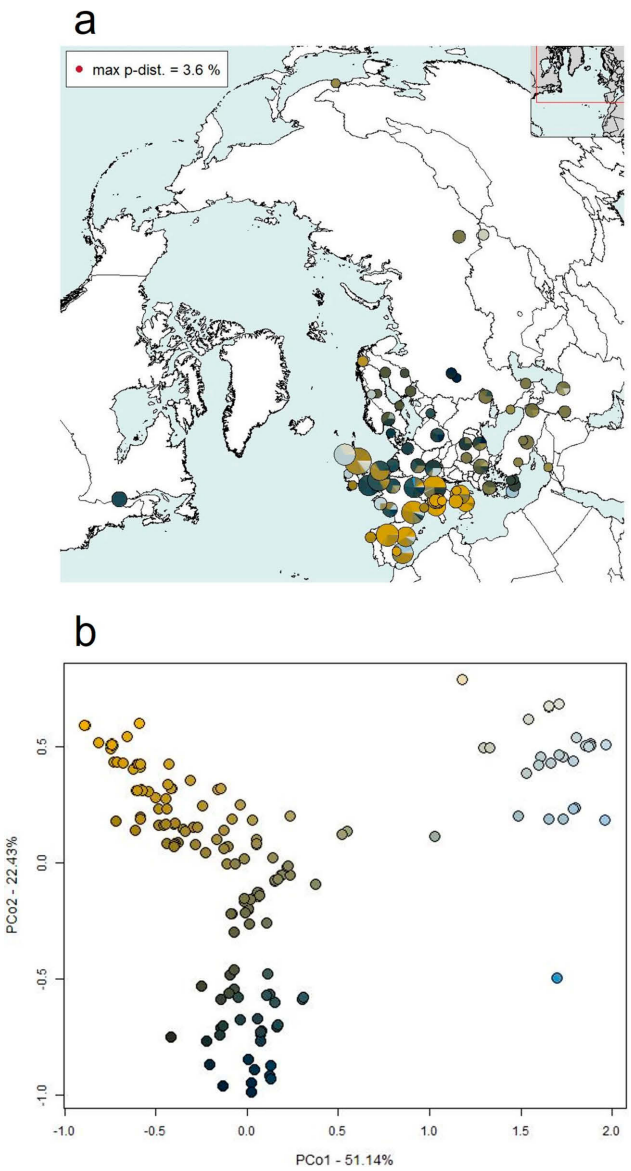

Fig. S2. Haplotype map for 611 specimens of *Polyommatus icarus* (a). Colors match the bidimensional color space of the PCoA projection of p-distances (b).

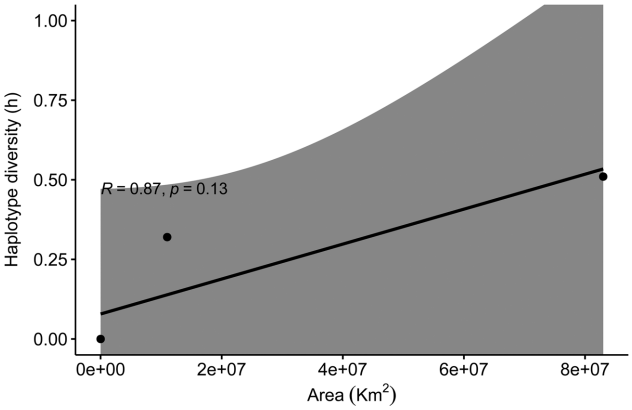

Fig. S3. Pearson's correlation between haplotype diversity (h) and area (km^2); the shaded area depicts 95% confidence interval.

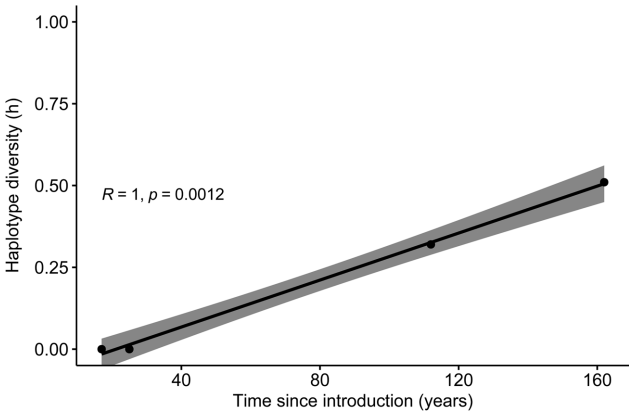

Fig. S4. Pearson's correlation between haplotype diversity (h) and time since introduction (years); the shaded area depicts 95% confidence interval.