ORIGINAL ARTICLE

A complete COI library of Samoan butterflies reveals layers of endemic diversity on oceanic islands

Claudia Bruschini[1] ⬤ | Eric D. Edwards[2] ⬤ | Gerard Talavera[3] ⬤ | Varea D. Vaurasi[4] |
Galumalemana F. Latu[4] | Leonardo Dapporto[1] ⬤

[1]ZEN Lab, Dipartimento di Biologia, Università degli Studi di Firenze, Sesto Fiorentino, Italia

[2]Department of Conservation, Conservation House, Wellington, New Zealand

[3]Institut Botànic de Barcelona (IBB), CSIC-Ajuntament de Barcelona, Barcelona, Spain

[4]Faculty of Science, National University of Samoa, To'omatagi, Samoa

Correspondence

Claudia Bruschini, ZEN Lab, Dipartimento di Biologia, Università di Firenze, Via Madonna del Piano 6, 50019, Sesto Fiorentino (FI), Italy.
Email: claudia.bruschini@unifi.it

Abstract

We investigated the entire butterfly fauna of the Samoan Archipelago (Pacific Ocean) by combining COI barcode sequences for specimens from these islands with those available in repositories at larger biogeographic scale. Haplotype networks and a generalized mixed Yule-coalescent (GMYC) model were applied to identify evolutionary significant units (ESUs). The ESUs from Samoan islands were compared with ESUs of the same or sister taxa regionally and worldwide to explore the level of endemicity and of congruence between established taxonomy and COI barcodes. The level of ESUs endemicity was similar to that shown by species and subspecies. Australia was the most frequent origin for Samoan lineages, followed by Orient-Asia. When comparing the agreement and mismatch between established taxonomy and ESUs between the Australia-Oceania region and Europe and North America, the COI molecular marker revealed a similar performance in taxonomic identification. Despite this overall convergent pattern, the degree of mtDNA divergence and the analysis of functional traits suggested that the mechanisms producing patterns of genetic differentiation in temperate butterflies over ancient continental lands differ to those occurring across a vast ocean into geologically young islands. Mechanisms on Samoan islands include relatively recent and exceptional oceanic dispersal, possibly followed by repeated extinction events. In the Australia-Oceania region we found a similar fraction of species showing introgression with the maintenance of phenotypic differences as it occurs on the mainland, but the phenomenon was limited to sectors of each species distribution area. Regular gene flow among the Samoan islands seems to prevent allopatric speciation within the archipelago.

KEYWORDS

biogeography, butterflies, continental fauna, endemics, haplotype network, mtDNA, oceanic islands, Samoan Archipelago

---

1 | INTRODUCTION

Islands have always played a key role in building evolutionary theory (cf. Whittaker et al., 2017). As discrete, manageable and replicable systems, islands represent excellent settings to study speciation and community assembly processes (Warren et al., 2015), given that their isolated spatial boundaries allow simplifying assumptions for experimental design and data interpretation (Vellend & Orrock, 2010; Whittaker & Fernández-Palacios, 2007). Islands contribute to global biodiversity in a higher proportion when compared to mainland areas of similar size (Whittaker & Fernández-Palacios, 2007). Oceanic islands comprise only about 5% of the Earth's land surface area, yet they have long captivated biogeographers for their splendid isolation and their unusual, sometimes bizarre, flora and fauna (Gillespie, 2001).

Molecular surveys of different insect groups on islands have also shown evolutionary determinism, species characterized by lower dispersal and a lesser degree of ecological generalism, have a higher probability of showing diverging populations and endemic elements (Dapporto et al., 2017; Matos-Maraví et al., 2014; Núñez et al., 2020; Salces-Castellano et al., 2021; Scalercio et al., 2020; Sourakov & Zakharov, 2011). Another common result of molecular surveys is that classic taxonomy lists based on morphological features show common cases of incongruence with assessments based on molecular approaches. A first phenomenon involves the existence of groups of taxa highly diversified on a genetic basis showing almost identical morphologies to each other (cryptic taxa, Fišer et al., 2018). It is well recognized that morphologically cryptic taxa represent a significant proportion of global diversity (Bickford et al., 2007; Adams et al., 2014). On the other hand, many species that are clearly ecologically and morphologically diversified can show a very similar genetic makeup with closely related taxa (Struck et al., 2018). This phenomenon can be due to the fast emergence of species-specific features (e.g. paraphyletic speciation), to post-speciation events of genetic introgression (Mallet, 2005) or to highly incomplete sampling of genomes.

Due to different speciation processes operating on oceanic islands compared to mainland (Whittaker et al., 2017; Whittaker & Fernández-Palacios, 2007), the patterns of discordance between the perceived diversity (morphology-based taxonomy) and genetic divergence could vary. Divergence usually occurs faster on islands after founder effects and genetic drift, and because of directional selection and radiation episodes following stepping-stone colonization to neighbouring islands (Clegg et al., 2002; Coyne & Orr, 2004; Lescak et al., 2015; Matos-Maraví et al., 2014; Núñez et al., 2020; Whittaker & Fernández-Palacios, 2007).

By contrast, in mainland scenarios, the geographical continuity of a population can be interrupted in time other than in space, mostly due to climatic changes, and lineages may speciate in parapatry involving episodes of gene flow (Mallet, 2005; Ebdon et al., 2021 for butterflies). Such mechanisms can produce introgression and genetic sweeps, notably in mtDNA, which may lead to lumped taxa in molecular-based species delimitation methods (D'Ercole et al., 2021; Dincă et al., 2015). This, in theory, should show a lower incidence on oceanic islands due to the virtual absence of gene flow between islands.

The evaluation of the relative incidence of these phenomena can only be achieved by applying a comparative approach, which involves species of a relatively large taxon (i.e. taxonomic coverage) and over an entire community (i.e. geographical coverage). Up to now, comparative genetic studies are limited by the lack of such complete genetic assessments and by the costs associated to obtain them. In recent decades a portion of the cytochrome *c* oxidase I mitochondrial DNA (COI) has gained a consensual and wide recognition as a barcoding marker complementing taxonomic practices and facilitating species identification (Hebert et al., 2003). Barcoding approaches to catalogue biodiversity have the advantage to rapidly produce extensive datasets to be used in comparative studies on diversity encompassed by mitochondrial genetic variation (Dapporto et al., 2019; Leigh et al., 2021; Theodoridis et al., 2021).

Butterflies are the insects that have received the strongest effort on the DNA barcoding of some entire continents (the western Palearctic Region, Dincă et al., 2015, Dapporto et al., 2022; North America, D'Ercole et al., 2021) and now repositories such as BOLD and GenBank contain hundreds of thousands of sequences to be used for comparative approaches. At the wide (continental) scale, the correspondence between classic and molecular taxonomy has been attempted in European (299 species in Dincă et al., 2015) and North American butterflies (755 species in D'Ercole et al., 2021). In those studies, the comparison of consensus taxonomic lists with a genetic species delimitation approach (GMYC) revealed that only 55.9% and 64.7% of published accepted species, respectively, correspond to the COI-inferred evolutionary significant units. In detail, 27.7% and 16.9% of named species could potentially be involved in cryptic speciation, while 16.4% of European and 18.4% of North American butterflies show low COI divergence and were lumped by the GMYC approach (D'Ercole et al., 2021; Dincă et al., 2015). At a smaller spatial scale the congruence between molecular and traditional taxonomy is greater, usually higher than 90% (Hausmann et al., 2011; Litman et al., 2018) mostly due to the lower incidence of pairs of allopatric cryptic taxa (Vodă et al., 2015).

No extensive molecular investigation of geospatial evolution has been carried out for butterflies of a tropical island, and neither for a highly insular and relatively small oceanic island. Comprehensive biodiversity assessments of entire insular taxa and communities are not only fundamental to fill the Linnean (species identification) and the Wallacean (distribution) shortfalls, but also to understand the phenomena that have delineated the observed diversity for unique island communities (Dapporto et al., 2017; Salces-Castellano et al., 2020; Shaw, 2002; Sourakov & Zakharov, 2011; Struck et al., 2018; Vodă et al., 2016).

In this study, we obtained COI sequences for virtually all the species occurring in the oceanic Pacific Archipelago of Samoan and American Samoan islands (hereafter referred to as Samoan Is., Figure 1a). The butterfly fauna occurring on these islands comprises several endemic taxa and species typical of the closest enlarged land masses (Australia and Southeast Asia).

The archipelago is relatively young since the most ancient volcanic activity in American Samoa is dated ~1–1.8 Mya while on Savai'i (Samoa) it is dated ~5.3 Mya (Strak & Schellart, 2018). The young geological origins and a minimal island area, when compared to continents, makes Samoan Is. a suitable archipelago to explore whether evolutionary speciation mechanisms producing diversification on islands and mainland could function differently.

In this paper, we provide a complete COI assessment for the butterflies of the Samoan Archipelago (Samoa plus American Samoa) and compare them with those available in public repositories. We apply the methodological approach used to contrast mtDNA diversity and established taxonomy in European and North American datasets to the butterflies of Samoan Is. in comparison with populations from Australia and Oceania islands. In particular, we (i) identify the main directions of colonization and the endemic mtDNA lineages of the Samoan Archipelago (ii) test for an unbalanced frequency of endemic entities in dispersive or sedentary species (iii) explore the level of congruence between established taxonomy and mtDNA, (iv) compare the pattern of congruence between butterfly populations from Samoan Is. with populations of Australian and Oceanian regions, and extended to two mainland models recently published (North America, D'Ercole et al., 2021 and Europe, Dincă et al., 2015).

2 | MATERIALS AND METHODS

2.1 | Study area and butterfly endemics

The Samoan Archipelago consists of four islands (>40 km^2) divided between the Samoa islands (Savai'i, Upolu) in the west and American Samoa (Tutuila and Ta'u) in the east (Figure 1a). A scatter of smaller islands also occurs. Samoa is recognized regionally as retaining important sites and species for biodiversity conservation including three out of 60 priority sites within the Polynesia Micronesia Global Biodiversity Hotspot (Myers et al., 2000; Atherton et al., 2007 for hotspot classification).

The Samoan Archipelago hosts 30 species of butterflies, and many populations are endemic at different levels: species level, *Papilio godeffroyi* – Pepe ae/Samoan swallowtail, Samoan ranger, *Deudorix doris,* Samoan dart, *Hypolimnas errabunda* – Samoan eggfly, *Phalanta exulans* and subspecies level, *Oriens augustula alexina* – Samoan cornealian, *Tirumala hamata melittula* – Samoan blue tiger, *Tirumala hamata tutuilae* – American Samoan blue tiger, *Euploea algea schmeltzi* – Samoan crow, *Euploea lewinii bourkei* – common crow, *Melanitis leda hopkinsi* – Samoan evening brown (Edwards, 2010).

2.2 | Butterfly sampling

During the years 2015–2017, extensive field surveys to collect butterflies were conducted in three islands: Savai'i and Upolu – the two main islands of Samoa – and Tutuila – the main island of American Samoa. Surveys spanned from sea level to mountain tops. Collection in Samoa was carried out with the support of NUS (National University of Samoa) and MNRE (Ministry of Natural Resources and Environment of Samoa) staff. In American Samoa, surveys were conducted with the aid of Dr. Schmaedick (American Samoa Community College). Specimens not found during field trips and needed for the completion of the dataset were gathered from American Community College collections. Out of the 30 species reported for the archipelago we were able to collect at least one and up to five specimens per species of 27 species for a total of 190 specimens. The three missing species are considered extinct in the Archipelago (Patrick & Patrick, 2012). The sequences have been publicly released in the BOLD project titled SAMOANGENMAP (code: SAMOA).

2.3 | Barcoding analysis

We sequenced the COI marker using the Biodiversity Institute of Ontario (Canada) facilities, following standard protocols for DNA barcoding (deWaard et al., 2008) with an ABI 3730XL capillary sequencer (Applied Biosystems). Permits to export the specimens were obtained from MNRE. COI sequences for each of the 27 species found in the Samoan Is. were merged with the
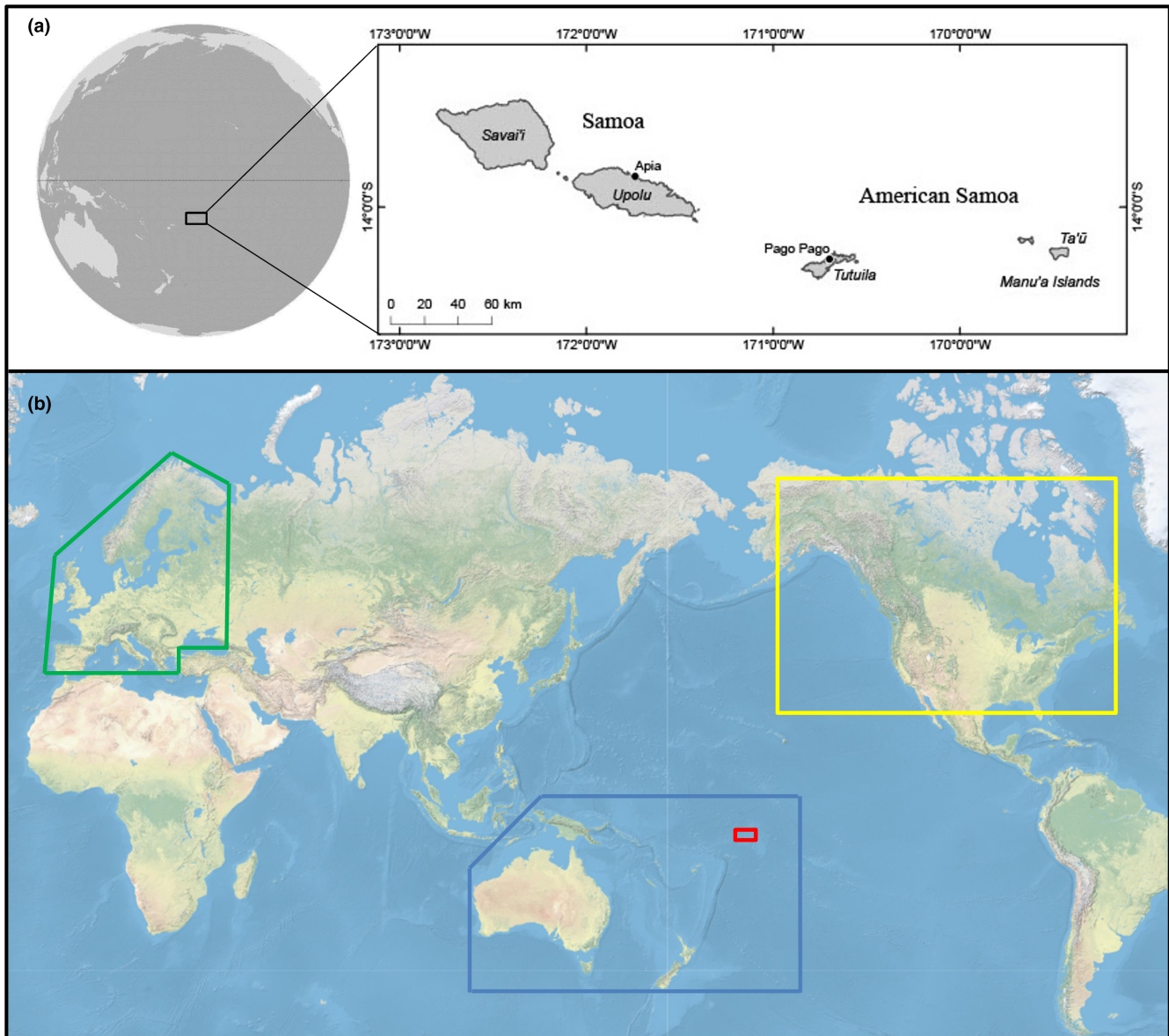
FIGURE 1 (a) Location of the Samoan Archipelago in the central South Pacific Ocean and (b) cross-regional comparisons of mtDNA libraries: the red box encompasses the Samoan Archipelago, the blue box includes the Australian and Oceanian regions for which the comparisons between classical and molecular taxonomy diversity patterns have been carried out for the Samoan butterflies species, the yellow and green boxes highlights the North American and European regions, respectively, where a similar comparative assessment was done.

sequences of public records available worldwide in the BOLD Data Systems and GenBank, creating a dataset with 901 DNA barcodes from 27 species belonging to 23 genera. Moreover, we collected all available world-wide sequences (5170) in BOLD and GenBank for all the species belonging to the genera occurring on Samoan Is. (global dataset). The fasta file containing BOLD and GenBank codes is available in the GitHub repository (https://github.com/leondap/files/blob/main/GMYC_data.fas).

2.4 | Phylogenetic analysis and GMYC delimitations

We collapsed the COI dataset for all studied genera at a global level to unique haplotypes using the 'haplotype' function of the R package 'pegas' (https://cran.r-project.org/web/packages/pegas/index.html). The number of haplotypes for the global dataset associated with the 23 Samoan genera was 2401 (894 within Nymphalidae, 364 within Pieridae, 7 within Hesperiidae, 829 within

Papilionidae and 307 within Lycaenidae-Riodinidae). We used BEAST 1.8 (Drummond et al., 2012) to reconstruct an ultrametric phylogenetic tree (Figure S1, Table S1) for each family of butterflies. Two independent chains of 100 million generations were run in BEAST for each dataset. The substitution model was set to GTR + I + G with six gamma rate categories. A coalescent tree prior was set. Divergence times were estimated by applying a strict clock and a normal prior distribution centred on the mean between two widely used substitution rates of 1.5% uncorrected pairwise distance per million years (Quek et al., 2004) and 2.3% (Brower, 1994). Values were sampled every 10% of the run length and convergence was inspected in Tracer v.1.6 (http://tree.bio.ed.ac.uk/software/tracer/).

We applied the general mixed Yule-coalescent model (GMYC, Fujisawa & Barraclough, 2013) for each family tree to identify evolutionary significant units (ESUs) using the R package 'splits' (https://cran.r-project.org/web/packages/SplitSoftening/index.html) with default settings. We choose the GYMC approach as it was the model used by the two articles (Dincă et al., 2015; D'Ercole et al., 2021) with which we compared our results. The GMYC delimitations for each of the 27 pre-established species among Samoan Is. (Patrick & Patrick, 2012) were categorized in (1) 'single entity species (SE)': all haplotypes of a species belong to a single GMYC ESU, (2) 'multiple entity species (ME)': haplotypes belong to two or more ESUs, (3) 'lumped entities (LE)': two or more species are recovered as a single ESU and (4) 'lumped + multiple entities (LME)': species are split in multiple ESUs and lumped with other species.

The delimitation category of a given taxon (and thus their regional incidence) may depend on the size of the region where it has been studied. In general, the same species can appear as a single entity if the assessment is made at a narrow geographic scale and as a multiple or lumped entity for wider geographic sampling areas (Talavera et al., 2013). We assessed the delimitation categories for the Samoan butterflies at two different scales: worldwide (W in Table 1) and Australian and Oceanian regions (AO in Table 1) as described by Holt et al. (2013) (Australia, New Zealand, Samoan Is., Polynesia, Papua New Guinea-Melanesia and Micronesia). The assessment of the frequency of delimitation categories at the wide continental scale, allowed us to make a direct comparison with the studies of North America and Europe.

The frequencies of butterfly species attributed to the four delimitation categories for Europe (Dincă et al., 2015) and North America (D'Ercole et al., 2021) were compared by Chi-squared tests (corrected for multiple comparisons using Bonferroni correction) with the results obtained for Samoan species when considering specimens collected in the Australian - Oceanian region. Endemic ESUs and haplogroups (H) of Australian-Oceanian, Oceanian and Samoan areas were also identified.

2.5 | Haplotype networks

Intraspecific relationships among haplotypes were built by haplotype networks using TCS 1.21 (Clement et al., 2000) and graphically improved using tcsBU (Múrias Dos Santos et al., 2016). We also created haplotype networks comprising more than one species, for lumped species and for endemic taxa showing low divergence in the phylogenetic BEAST tree from the global dataset. To explore colonization events, we visually inspected the distribution of haplotypes and look for consistency with the global zoogeographic regionalization proposed by Holt et al. (2013).

2.6 | Functional traits

Among possible functional traits, we selected two that are biologically meaningful in the context of dispersal capacity and speciation in isolated island archipelagos: wingspan and migratory/non-migratory behaviour. Wingspan is largely used in butterflies as a proxy for dispersal capability (Sekar, 2012), and it shows negative correlation with genetic diversification (Dapporto et al., 2017, 2019; Scalercio et al., 2020). We retrieved the wingspan for each species (Table 1) as the mean between maximum and minimum wingspan recorded by Patrick and Patrick (2012). Although other alternative measures have been identified as better predictors for mobility like wing ratio and wing loading (Dudley & Srygley, 1994; Habel et al., 2018), they are not available for Samoan butterflies. Information about migratory behaviour of Samoan butterflies was obtained following criteria and literature gathered in García-Berro et al. (2022) and Chowdhury et al. (2021).

Our goal was to assess differences in wingspan and in frequency of migratory species between species belonging to different endemic status and delimitation categories. Butterfly species from different families largely differ in their wingspan making assessment of the entire community prone to biases due to unbalanced composition of the groups. To standardize wingspan, we scaled wingspan for the three families comprising more than 2 species (Lycaenidae, Pieridae, Nymphalidae) to the same mean (0) and standard deviation (1) using the *scale* R function (base package). Then, we compared wingspan by using Wilcoxon signed rank test (stats package for R) between

TABLE 1 Outcomes of barcoding and GMYC delimitations for the Samoan Is. butterfly fauna.

Species	WS	End	ESUs	W	AO	End ESU AO	End ESU O	End ESU Samoan is.	End H AO	End H O	End H Samoan is.	% div	Mig
Badamia exclamationis	51		2	ME	SE	Yes	No	no	yes	no	no		yes
Papilio godeffroyii	96.5	sp	1	SE	SE	Yes	yes	yes	yes	yes	yes	6.4	
Appias athama	59		1	SE	SE	yes	yes	?	yes	yes	?	?	
Appias paulina	55		2	ME	SE	yes	no	no	yes	yes	yes	0.2	yes
Belenois java	55		1	SE	SE	yes	no	no	yes	yes	yes	0.3	yes
Eurema hecabe	34		2	ME	ME	yes	yes	yes	yes	yes	yes	2.0	yes
Catochrysops taitensis	20.5		1	LME	LME	yes	no	no	yes	yes	yes	0.3	
Lampides boeticus	23.5		4	ME	ME	no	no	no	no	no	no	–	yes
Zizina otis	23		1	LU	SE	no	no	no	yes	no	no	–	yes
Zizula hylax	18		2	ME	SE	yes	no	no	yes	no	no	–	
Petrelaea tombugensis	21.5		1	LE	LE	no	no	no	no?	no?	no	–	
Jamides argentina	23.5	sp	1	SE	SE	yes	yes	yes	yes	yes	yes	1.4	
Euchrysops cnejus	25.5		1	SE	SE	no	no	no	no	no	no	–	
Nacaduba dyopa	22		2	ME	ME	yes	yes	?	yes	yes	?	?	
Deudorix doris	33	sp	1	SE	SE	yes	yes	yes	yes	yes	yes	2.6	
Danaus plexippus	86		1	SE	SE	no	no	no	no	no	no	–	yes
Tirumala hamata	74	ssp	1	LE	LE	no	no	no	yes	yes	yes	0.5	yes
Euploea lewinii	67	ssp	1	LE	LE	?	?	no	?	?	no	–	
Euploea algea	59	ssp	1	LE	LE	?	?	no	?	?	?	?	
Melanitis leda	66	ssp	3	ME	ME	yes	yes	no	yes	yes	no	–	yes
Doleschallia tongana	63		1	SE	SE	yes	yes	?	yes	yes	?	?	
Hypolimnas bolina	73		7	ME	ME	yes	yes	no	yes	yes	no	–	yes
Hypolimnas antilope	64		1	LE	LE?	no	no	no	yes	no	no	–	
Hypolimnas erabunda	73	sp	1	SE	SE	yes	yes	yes	yes	yes	yes	1.4	
Junonia villida	42		1	SE	SE	?	no	no	?	no	no	–	yes
Phalanta exulans	38.5	sp	1	SE	SE	yes	yes	yes	yes	yes	yes	5.0	yes
Vagrans egista	52.5		4	ME	SE	yes	no	no	yes	yes	yes	0.6	yes

Abbreviations: % div, percentage of mtDNA divergence; AO, Australian and Oceanian regions; End, endemic; ESUs, Evolutionary Significant Units; H, haplotype; Mig, Migratory behaviour; O, Oceanian regions; W, worldwide; WS, wingspan (mm) for each species.
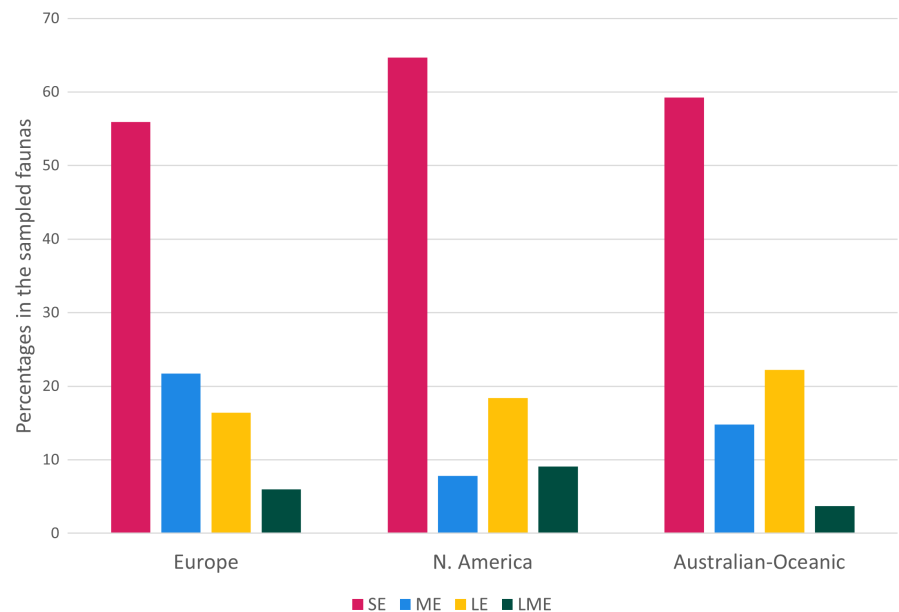
species showing endemic and non-endemic entities (species/subspecies, ESUs, haplotypes) and by Kruskal-Wallis test (stats package for R) among the three categories of SE, LE, ME (except for LME represented by a single species). Similarly, Fisher's exact tests were used to compare the frequency of endemic entities (species/subspecies, ESUs, haplotypes) and of the three categories (SE, LE, ME) between migratory and non-migratory species.

3 | RESULTS

3.1 | Global dataset

We found 403 GMYC ESUs in the global dataset for the 23 genera sampled in the Samoan Is. The divergence threshold identified by GMYC for the identification of ESUs was 2.07 My. When inspecting GMYC delimitation categories of specimens collected worldwide, 11 Samoan species out of 27 species showed a single ESU per species (SE), 6 species were lumped (LE) with other species, 9 were divided in multiple ESUs (ME) and one showed both lumped and multiple entities (LME, *Catochrysops taitensis*, which showed two entities but one of them was lumped with *C. panormus*) (Table 1). When the assessment was restricted to specimens belonging to the Australian continent and Oceanian regions the frequencies were: 16 SE, 5 LE, 5 ME, 1 LME (Figure 2, Table 1). These frequencies did not show significant differences compared to those found for the largest datasets including European and North American butterflies (chi-squared = 4.667, p = .198 N = 326 with Europe and chi-squared = 2.767, p = .429, N = 782 with North America) (Figure 2).

3.2 | Haplotype networks for Samoan taxa

The haplotype networks could only be calculated for 24 of the 27 Samoan species collected as specimens of 3 species were not available in BOLD or GenBank from other regions. Each Samoan species show that 17 of 24 species were endemic ESUs in the entire Australian-Oceanian region. Among them, 11 were also endemic to the Oceanian regions (Fiji, New Caledonia, French Polynesia) and 6 (25% of species) revealed to be endemic to Samoan Is. (Table 1). Haplotype networks (Figures 3,4,5) also revealed several endemic monophyletic groups of closely related haplotypes (occurring in 20 species at Australian-Oceanian level, 16 at Oceanian level and 11 for Samoan Is.). This suggests that Samoan butterflies might be the result of radiations centred on the older landmasses of Australia and Oceania (e.g. Papua New Guinea, New Caledonia and others). We could also identify significant incongruence with established taxonomy. Among the unexpected endemic ESUs we found one haplotype of *Eurema hecabe* separated by a minimum of 13 mutations (1.98% of COI divergence). Other very distinctive ESUs were found in *Appias paulina*, and less markedly in *Vagrans egista* (Figure 4).

3.3 | Functional traits analyses

Samoan endemic entities (species/subspecies, GYMC ESUs and haplotypes) did not show significant differences in wingspan (scaled among families) compared to non-endemic entities (Wilcoxon rank sign test, W = 50, p = .3077; W = 45, p = .8754 and W = 57, p = .9158,

FIGURE 2 Percentages of butterfly species by four delimitation categories according to the GMYC outcome. SE, single entities; ME, multiple entities; LE, lumped entities; LME species showing both lumped and multiple entities. Percentages are comparatively shown between Europe, North America and Samoan species for the assessment at the Australian-Oceanic scale.
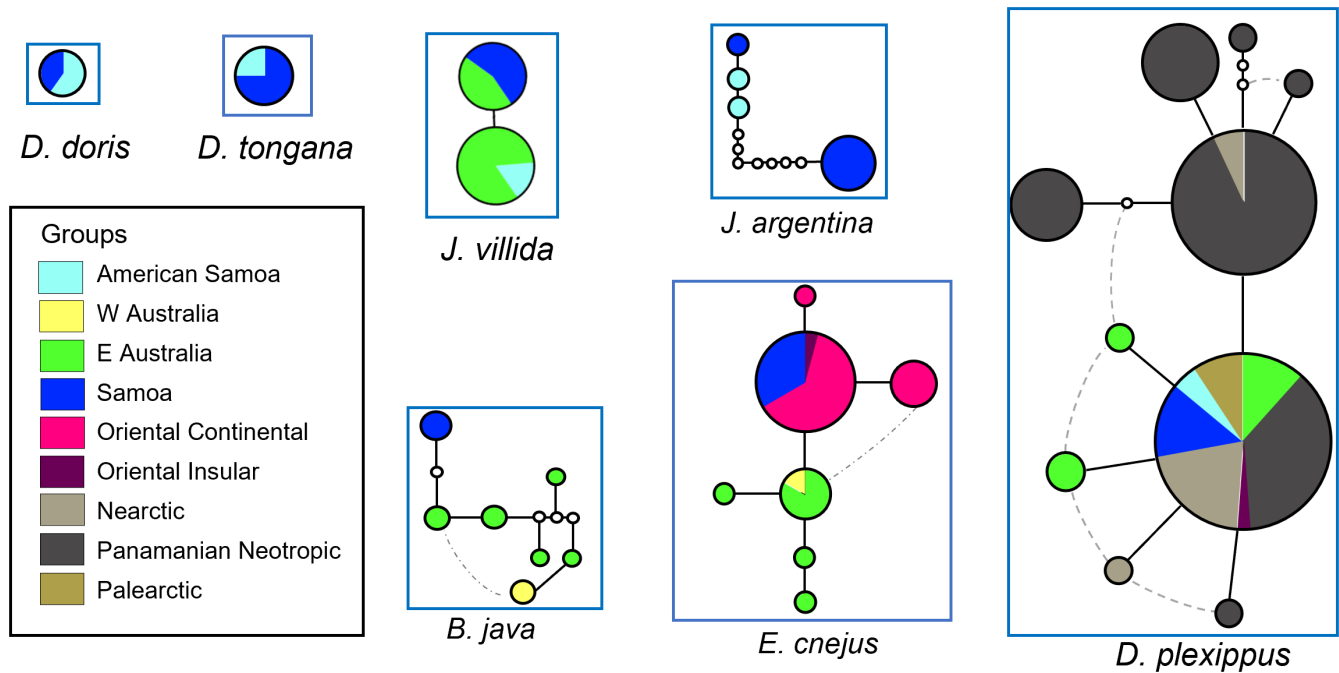
FIGURE 3 Haplotype networks of six species showing a single entity (SE) from the GMYC outcome. *P. godeffroyii, H. errabunda* and *P. exulans* are not shown as they are endemism of Samoa or American Samoan and no specimens of the three species could be found in BOLD or GenBank for other regions. Colours within the pies indicate geographic regions, following Holt et al. (2013). Each segment represents one mutation, white circles represent missing haplotypes, and the size of the circles is proportional to the number of specimens represented in each haplotype. Loops are represented with dotted grey lines. The legend indicates the geographical location of each sequence. Blue frames indicate the presence of only one ESU as determined by GYMC. Full scientific names are provided in Table 1.

Figure 6a, respectively). Moreover, no significant differences in wingspan were found among species showing different types of concordance/discordance between taxonomic and GMYC assessments (Kruskal-Wallis test, chi-squared = 1.615, $df = 2$, $p = .445$, Figure 6b).

Non-migratory species show a higher frequency of endemic taxa than migratory species (Fisher's exact test, $p = .041$) while frequencies were similar once considering ESUs and haplotypes (Fisher's exact test, $p = .710$; $p = .166$, respectively).

No significant differences were found when comparing frequencies of SE, LE, ME categories between migratory and non-migratory taxa (Fisher's exact test, $p = .212$).

4 | DISCUSSION

Although the use of a single mitochondrial marker is not expected to be decisive for taxonomy (Dasmahapatra et al., 2010; Shaw, 2002; Zamani et al., 2022), the increasing availability of COI sequences has unveiled hidden dimensions of diversity, thus complementing taxonomic assessments (Galimberti et al., 2021; Hebert, Ratnasingham, & DeWaard, 2003; Matos-Maraví et al., 2014; Núñez et al., 2020; Wiemers et al., 2018). The COI-based survey of the Samoan butterflies reveals that the ratio of species

endemicity (as per the accepted taxonomy) is similar to that resulting from mitochondrial DNA haplotypes. However, we highlight several incongruences between taxonomic assessments and COI differentiation. For example, in the form of different taxonomic species lumped in a single GMYC entity or for highly diverging populations showing no morphological characterization. The frequency of endemic haplotypes and incongruent identifications between GMYC delimitation and established taxonomy was not associated to wing size or migratory behaviour, traits known to correlate with species dispersal capacity and genetic divergence (Dapporto et al., 2017; Scalercio et al., 2020; García-Berro et al., 2022).

4.1 | Mitochondrial and taxonomic endemicity for Samoan islands and direction of natural colonization

The Samoan Archipelago shows a remarkable number of endemic butterfly taxa with 11 species or subspecies representing 33% endemism. Taxonomic endemicity is mirrored by the high fraction of endemic haplotypes occurring in at least 11 species. However, among the 11 species showing endemic haplotypes, four do not belong to the 11 endemic species/subspecies, which could turn into
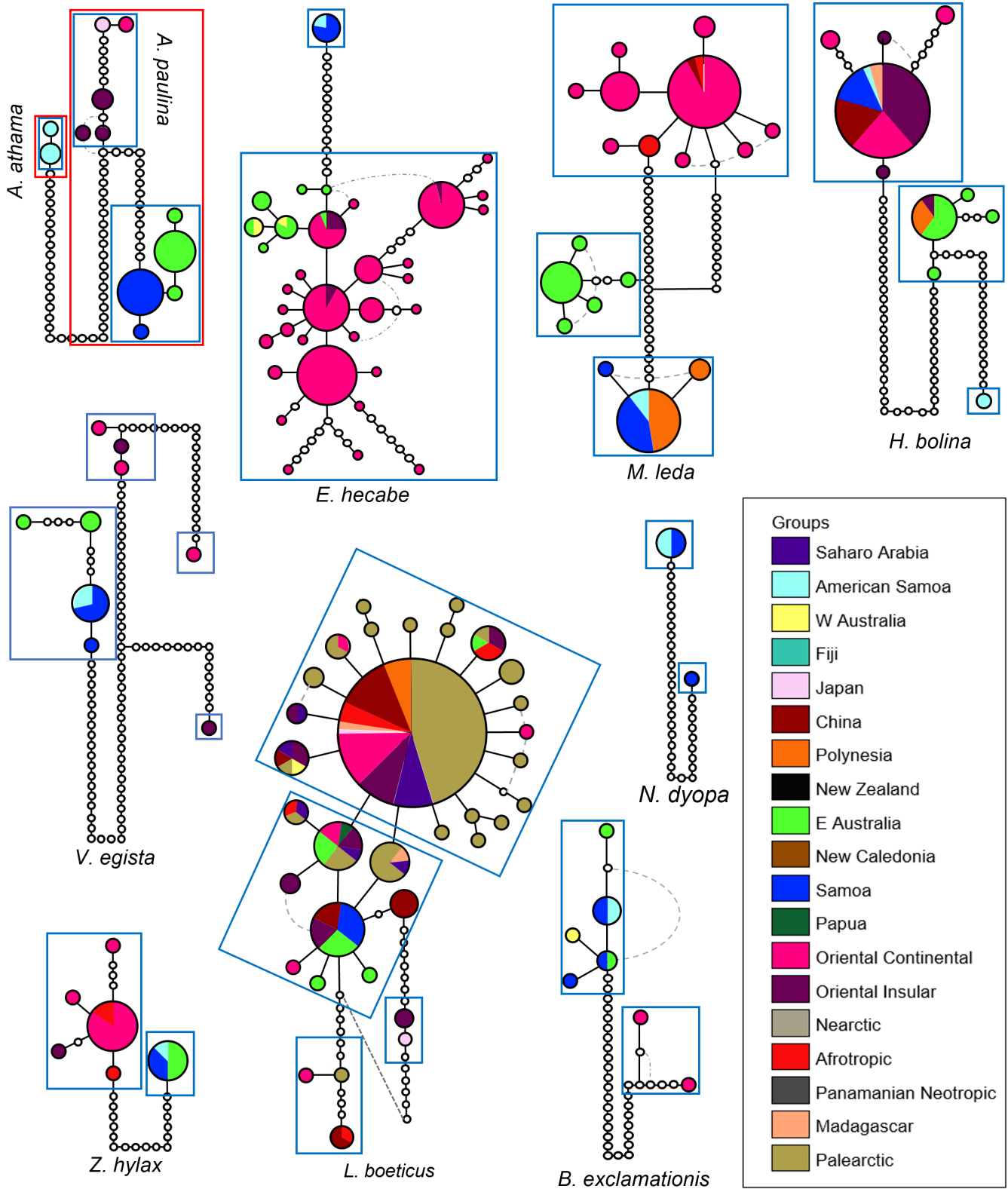
FIGURE 4 Haplotype networks of the species showing multiple entities (ME) from the GMYC outcome. Colours within the pies indicate geographic regions, following Holt et al. (2013). Each segment represents one mutation, white circles represent missing haplotypes and the size of the circles is proportional to the number of specimens represented in each haplotype. Loops are represented with dotted grey lines. The legend indicates the geographical location of each sequence. Blue frames indicate the presence of only one ESU as determined by GYMC. The red rectangles include specimens attributed to the same taxonomic species in case of discordant patterns. Full scientific names are provided in Table 1.
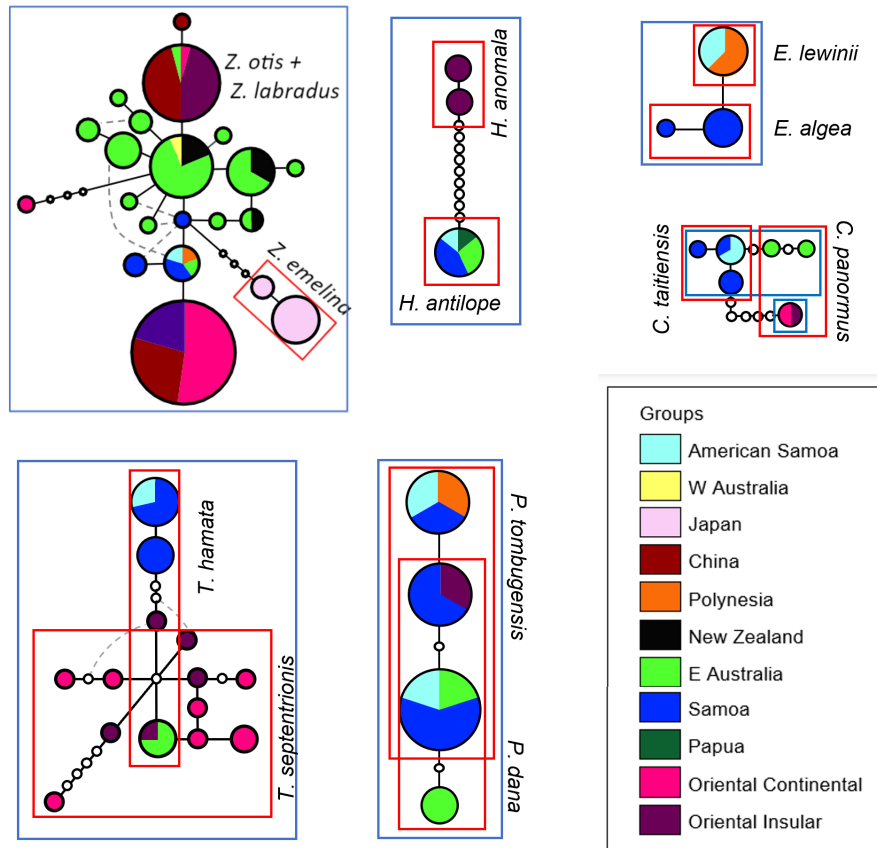
FIGURE 5 Haplotype network of the species showing lumped entities (LE) from the GMYC outcome. Due to the taxonomic uncertainty and potential for inaccurate identification of sequenced specimens, we included *Zizina otis* and *Z. labradus* sequences in the same blue frame. Colours within the pies indicate geographic regions, following Holt et al. (2013). Each segment represents one mutation, white circles represent missing haplotypes, and the size of the circles is proportional to the number of specimens represented in each haplotype. Loops are represented with dotted grey lines. The legend indicates the geographical location of each sequence. Blue frames indicate the presence of only one ESU as determined by GYMC. The red rectangles include specimens attributed to the same taxonomic species in case of discordant patterns. Full scientific names are provided in Table 1.

an even higher number of endemic taxa (15) if a combined approach (established taxonomic taxa complemented by gene information) is applied (Menchetti et al., 2021).

A major difference between the populations living in the Samoan Archipelago compared to those from Europe is the lack of correlation between genetic differentiation and functional traits determining dispersal (and thus gene flow). In Europe (Dapporto et al., 2019) and on Mediterranean fragmented islands (Vodă et al., 2016; Dapporto et al., 2017; Scalercio et al., 2020), smaller butterfly species and those with shorter flight periods, are characterized by less differentiation and greater endemicity, while almost none of the migratory species show endemic lineages. On the Samoan Archipelago, six species of the 12 species reported as migratory, possess endemic haplotypes and the larger species in each family did not show a higher fraction of endemicity compared to the smaller ones. This result can be partly due to lower number of examined species, but it also reinforces the assumption that the great isolation of this archipelago represents an almost

insurmountable barrier to butterflies and that even highly dispersive species have colonized these islands only after exceptional and stochastic events.

In oceanic archipelagos, many endemic taxa show a concordance between island age and lineage ages, where younger islands are inhabited by younger lineages (progression rule, Shaw & Gillespie, 2016). This does not occur in our system since five of the six species showing endemic haplotypes in western Samoa (Upolu and Savai'i) and American Samoa (Tutuila) shared their haplotypes. Thus, most endemics within the archipelago occur in multiple islands. This can be due to ongoing intra-archipelago gene flow swamping inter-island diversification.

Human assisted colonization may have led to the establishment of undifferentiated populations. For example, *B. exclamationis* (Hesperiidae) and *P. tombugensis* (Lycaenidae) both feed on almond trees (Combretaceae; *Terminalia* spp.). Their host plant *Terminalia catappa* is possibly native but also known to be widely dispersed among Pacific islands by Polynesian voyaging (Thomson
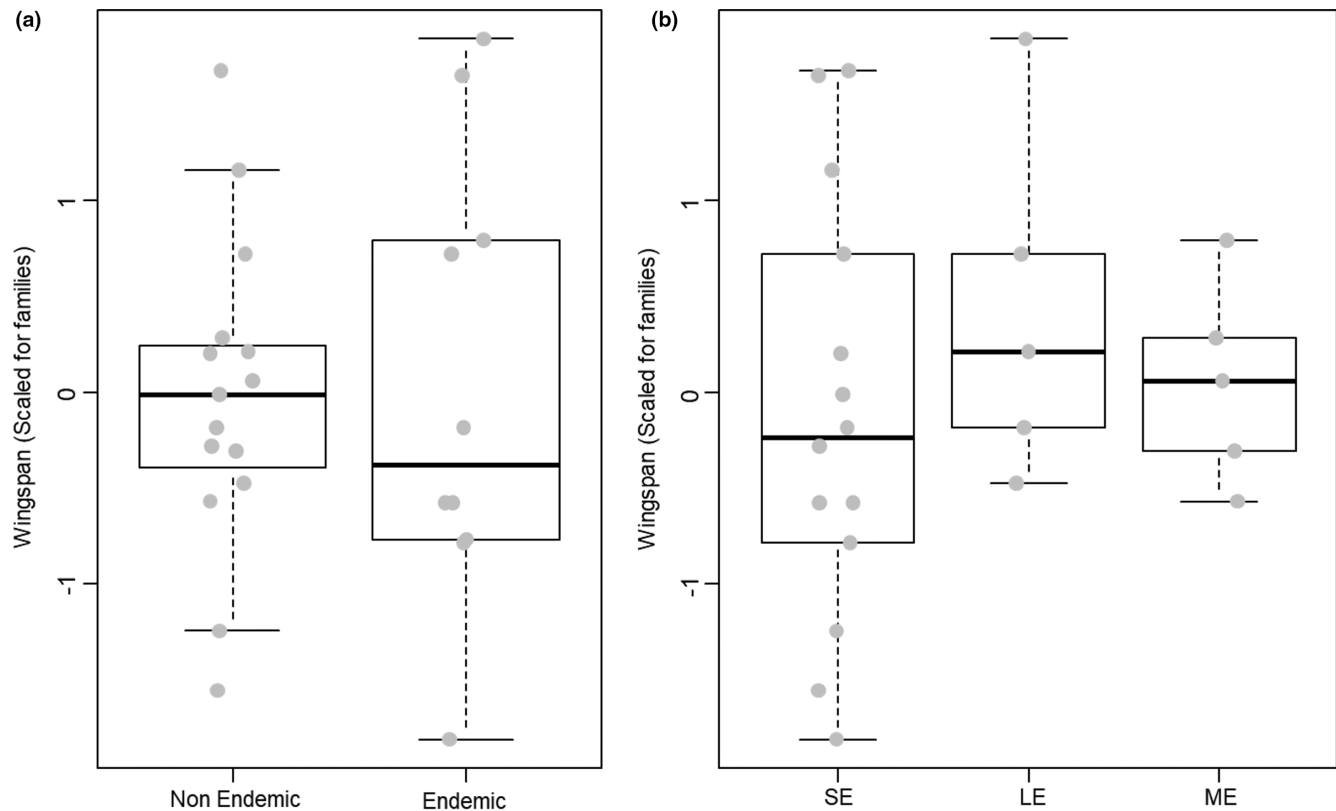
FIGURE 6 A comparison of butterfly wingspan (scaled millimetres) between species tested for (a) endemic and non-endemic haplotypes in Samoan Is. and, (b) contrasted among GMYC delimited ESUs showing single entities (SE), lumped entities (LE) and multiple entities (ME).

& Evans, 2006) and this may have aided butterfly dispersal directly and/or by providing a suitable host for establishment. In support of this possibility, neither species has endemic haplotypes. Two other butterflies have essentially followed the human establishment of their host plants. Both *D. plexippus* and *D. tongana* colonized after caterpillar hosts in Apocynaceae and Acanthaceae, respectively, were established. Our molecular evidence supports the hypothesis for *D. plexippus* but lacks data from other parts of the Pacific to interpret for *D. tongana*.

According to Strak and Schellart (2018), Western Samoa islands emerged about 5.3 My ago. By using evaluation of insect mutation rates, a maximum divergence originated on Western Samoa (2.8 My old) is expected to vary between 1.5%/My*2.8My = 4.2% (Quek et al., 2004) and 2.3%/My*2.8My = 6.4% (Brower, 1994). This range encompasses the minimum COI p-distance for the known closest relative of all the Samoan endemic species and lineages, comprising the most divergent *Papilio godeffroyi* (6.4% from *P. schmeltzi* nearest sister occurring on Fiji Islands). A similar reasoning applies to American Samoa where the oldest volcanic activity is dated between 1–1.8 My and for which a maximum mtDNA divergence evolved in situ should range between 2.1% and 3.2% (considering 1.4 My, the average between 1.0 and 1.4). This

hypothesis excludes the formation on this island of only *Papilio godeffroyi* and *Phalanta exulans* showing a divergence of 6.4 and 5.0%, respectively (Table 1). So, we can conclude that most endemics could have originated in either island and then colonized the other(s).

At the taxonomic level, the concordance among island endemics is much lower. Indeed, the two island systems only share two endemic species out of 11 (*J. argentina* and *D. doris*) and no subspecies. The faunistic differentiation could have been exacerbated by stochastic extinctions as envisaged by the island equilibrium theory (MacArthur & Wilson, 1967) or induced by relatively recent human activities as occurred for *P. godeffroyi* sixty years ago. This species was widespread and common in the entire archipelago but due to deforestation and likely other pressures, it has declined dramatically and now appears restricted to American Samoa ~5% of former range (Edwards, 2010; Patrick & Patrick, 2012).

At a wider scale, the haplotype networks indicate that most of the Samoan populations have their closest relatives in the Australian area. New Guinea may also be a source of many Samoan populations but the scarce availability in BOLD and GenBank of sequences for butterflies from this region did not allow us to thoroughly investigate this colonization pathway.

4.2 | Level of congruence between established taxonomy and DNA barcoding

When the sequences of the butterflies from the Samoan Is. were pooled with those from Australia and Oceania, the relative frequency of incongruences between taxonomy and GYMC classification did not differ from that found for European or North American butterflies. Moreover, the frequency of endemic haplotypes in Samoa Is. (41%) was similar to the one found in the island of Sicily, adjacent to continental Europe (37%, Scalercio et al., 2020). This occurred despite the diversification processes, operating in small and geologically young oceanic island communities, which are expected to differ from processes occurring over wide and ancient mainland regions. Based on our data, it is possible to establish whether different evolutionary processes have produced a convergent pattern, or if in contrast, similar mechanisms are perhaps occurring on oceanic and mainland populations. The most striking result is the absence of a higher fraction of ME in Samoan Is. species compared to mainland. Most of the ME taxa highlighted here are represented by groups of populations from different islands recognized as conspecific and separated by a 2%–3% of COI divergence. It is well known that GMYC can often produce different entities among conspecific populations, which have undergone differentiation in allopatry (Sukumaran & Knowles, 2017). The geologic age of the Samoan Archipelago (5.3 My for the oldest island) is not much higher than the threshold identified by GMYC for the transition between intraspecific to interspecific genetic diversification (2.07 My). It is thus possible that the time since island emergence has not been sufficient to produce a high number of endemic GMYC ESUs. Moreover, based on the island equilibrium theory (MacArthur & Wilson, 1967; Whittaker et al., 2017), island populations, mostly in remote and small island systems, may undergo colonization and extinction events, which can shorten their persistence on islands over long geological times. This is potentially visible in five species among the 27 species examined where endemic haplotypes did not diverge at the GYMC threshold: *A. paulina*, *B. java*, *C. taitiensis, T. hamata* and *V. egista* (see Figures 3,4,5). If most of the colonizations of the Samoan islands have occurred after exceptional dispersal events as supposed above, the relatively small size of the islands and the absence of a rescue effect may have increased the risk for local extinction. This may have lowered the possibility that an endemic population persisted for a very long time, thus producing a majority of slightly diverging endemic haplotypes. In this perspective, the recent island age and the great isolation can determine the observed convergence in ME frequency with respect to older and wider mainland regions.

There is increasing evidence that mitochondrial diversification can be driven and maintained by maternally inherited infection by microorganisms interfering with insects' reproductive systems (notably *Wolbachia*). *Wolbachia* is known to play a major role in spatial structuring of mtDNA in both mainland and island butterfly populations and, notably, the populations of *H. bolina* from Oceanic islands comprising two concordant lineages of *Wolbachia* and mtDNA is one of the most studied examples for this phenomenon (Arif et al., 2021; Gaunet et al., 2019; Reynolds et al., 2019; Ritter et al., 2013; Sahoo et al., 2018; Sucháčková Bartoňová et al., 2021). A complete survey of *Wolbachia* occurrence and strain identification for the strongly diverging Samoan clades could identify the importance of this phenomenon, together with island age and isolation, in shaping mtDNA diversity across Oceanic islands.

4.3 | Interpreting the high proportion of lumped entities

Surprisingly, LE were more frequent than ME among Samoan butterflies, with an incidence similar to that found on mainland regions. Taxonomic species can be lumped in a species delimitation analysis based on mtDNA because of two main processes: (1) the described morphological variation may not be reflected by mtDNA variation (Struck et al., 2018), being low and below the GMYC threshold. In this case, although GMYC is not highlighting two different taxa, they can still be identified as different haplogroups (Dincă et al., 2021).

(2) two species differentiated in allopatry for morphological and genetic characters can get in secondary sympatry or parapatry over mainland or following exceptional dispersal events among islands (Cong et al., 2017; Mallet, 2005). In contact areas, hybridization is possible and mtDNA can be passed from one species to another by adaptive introgression (Mallet, 2005; Cong et al., 2017). The resulting genetic sweeps can erase the genetic diversity at the mitochondrial level while preserving species integrity at nuclear loci, determining the maintenance of phenotypic differences (Toews & Brelsford, 2012; Cong et al., 2017). Over Europe and North America both processes are common and about 15% of butterfly species share COI sequences with at least another species (Dincă et al., 2021; D'Ercole et al., 2021).

Low genetic diversification below the GMYC threshold due to short divergent time is likely to happen on relatively young oceanic islands. Possible examples of species pairs occurring in Samoa and in the Australian-Oceanic islands

include *E. algea-E. lewinii* and *H. antelope-H. anomala*, that were lumped by the GYMC approach. However, there are also examples of species sharing haplotypes (e.g. *T. hamata - T. septentrionis* and *C. taitiensis - C. panormus*) suggesting cases of introgression. Nevertheless, the occurrence of introgression showed a different pattern on islands, as our results document that isolated archipelagos can preserve their supposed ancestral populations from (adaptive) introgression. This pattern is evident in the Australia-Oceanian regions for *T. hamata-T. septentrionis* and *C. taitiensis - C. panormus* where the introgression appears incomplete and limited to some islands. On the other hand, in the absence of physical barriers, introgression usually proceeds across mainland regions until one of the two genotypes is completely erased (Cong et al., 2017). In line with some of our examples, introgression was stopped at a short sea strait in the Mediterranean Sea (Scalercio et al., 2020). As above, infection by different *Wolbachia* strains might also impede introgression (Arif et al., 2021).

It must be noted that the scarcity of COI barcodes of butterflies across the Oceanian, Australian and Oriental regions can hamper an exact location and date of parapatry, radiation-allopatry and introgression events. Our haplotype networks suggest that some of the diversification mechanisms observed over mainland areas could have occurred for the fauna of the remote Samoan Archipelago. A clear example of this possible limitation is *E. hecabe*, a recent human assisted colonizer of the Samoan Is., but that belongs to a highly diverging and apparently endemic clade for Samoa. *E. hecabe* has colonized Western Samoa over the past few decades (Tennent, 2006) and American Samoa no longer than a decade ago (Patrick & Patrick, 2012). In this view, the original population of this entity possibly belongs to the subspecies *E. hecabe sulphurata,* described from New Caledonia, Vanuatu and Fiji Islands, which has not been sequenced yet. However, this ESU appears to represent a discovered cryptic species rather than a subspecies.

5 | CONCLUSIONS

For 27 currently recognized butterfly species and subspecies of the Samoan Is., COI sequencing of the DNA barcode segment coupled with the GYMC approach has supported existing taxonomy for 11 species and uncovered hidden diversity among the others. Noteworthy, five butterflies considered endemic were all confirmed in our analysis, whilst 16 species from the 27 examined showed inconsistency between taxonomy and the GMYC molecular delimitations. From a taxonomic point of view, these inconsistencies open up the possibility of the existence of

overlooked cryptic species and species synonyms in the islands. To further assess this hypothesis additional support from independent characters (e.g. nuclear markers, genitalia morphology, ecological niche preferences) would be needed. Moreover, from an eco-evolutionary point of view, we showed that the frequency of taxonomic and genetic incongruency is similar between an island area and mainland regions despite the systems being characterized by very different ecological and historical settings. The comparison with island age and the absence of a correlation with dispersal capacity, indicate that most likely different mechanisms have produced a similar pattern. The increasing availability of DNA sequencing could greatly fuel comparative analyses and improve the understanding of the interplay between diversification processes and regional features.

ORCID

Claudia Bruschini 🔟 https://orcid.org/0000-0002-9203-7171
Eric D. Edwards 🔟 https://orcid.org/0000-0002-2211-7184
Gerard Talavera 🔟 https://orcid.org/0000-0003-1112-1345
Leonardo Dapporto 🔟 https://orcid.org/0000-0001-7129-4526

REFERENCES

Adams, M., Raadik, T. A., Burridge, C. P., & Georges, A. (2014). Global biodiversity assessment and hyper-cryptic species complexes: More than one species of elephant in the room? *Systematic Biology*, *63*(4), 518–533. https://doi.org/10.1093/sysbio/syu017

Arif, S., Gerth, M., Hone-Millard, W. G., Nunes, M. D. S., 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. *Molecular Ecology*, 30(20), 5196–5213. https://doi.org/10.1111/mec.16126

Atherton, J., Axford, J., Dowdeswell, N., Farley, L., James, R., Langhammer, P., Martel, F., Manner, H., Olson, D., & Sesega, S. (2007). Ecosystem profile. Polynesia-Micronesia biodiversity hotspot. In J. Atherton (Ed), *Conservation International-Melanesia Center for Biodiversity Conservation and Secretariat of the Pacific regional environment program*. Arlington.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K., & Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, 22(3), 148–155. https://doi.org/10.1016/j.tree.2006.11.004

Brower, A. V. Z. (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(14), 6491–6495. https://doi.org/10.1073/pnas.91.14.6491

Chowdhury, S., Fuller, R. A., Dingle, H., Chapman, J. W., & Zalucki, M. P. (2021). Migration in butterflies: A global overview. *Biological Reviews*, 96(4), 1462–1483. https://doi.org/10.1111/brv.12714

Clegg, S. M., Degnan, S. M., Moritz, C., Estoup, A., Kikkawa, J., & Owens, I. P. F. (2002). Microevolution in Island forms: The roles of drift and directional selection in morphological divergence of a passerine bird. *Evolution*, 56(10), 2090–2099.

Clement, M., Posada, D., & Crandall, K. A. (2000). TCS: A computer program to estimate gene genealogies. *Molecular Ecology*, 9(10), 1657–1659. https://doi.org/10.1046/j.1365-294X.2000.01020.x

Cong, Q., Shen, J., Borek, D., Robbins, R. K., Opler, P. A., Otwinowski, Z., & Grishin, N. V. (2017). When COI barcodes deceive: Complete genomes reveal introgression in hairstreaks. *Proceedings of the Royal Society B: Biological Sciences*, 284(1848), 20161735. https://doi.org/10.1098/rspb.2016.1735

Coyne, J. A., & Orr, A. H. (2004). Species: Reality and concepts. In *Speciation* (Vol. 37). Sinauer associates Sunderland, MA.

Dapporto, L., Cini, A., Menchetti, M., Vodă, R., Bonelli, S., Casacci, L. P., Dincă, V., Scalercio, S., Hinojosa, J. C., Biermann, H., Forbicioni, L., Mazzantini, U., Venturi, L., Zanichelli, F., Balletto, E., Shreeve, T. G., Dennis, R. L. H., & Vila, R. (2017). Rise and fall of Island butterfly diversity: Understanding genetic differentiation and extinction in a highly diverse archipelago. *Diversity and Distributions*, 23(10), 1169–1181. https://doi.org/10.1111/ddi.12610

Dapporto, L., Cini, A., Vodă, R., Dincă, V., Wiemers, M., Menchetti, M., Magini, G., Talavera, G., Shreeve, T., Bonelli, S., Casacci, L. P., Balletto, E., Scalercio, S., & Vila, R. (2019). Integrating three comprehensive data sets shows that mitochondrial DNA variation is linked to species traits and paleogeographic events in European butterflies. *Molecular Ecology Resources*, 19(6), 1623–1636. https://doi.org/10.1111/1755-0998.13059

Dapporto, L., Menchetti, M., Vodă, R., Corbella, C., Cuvelier, S., Djemadi, I., Gascoigne-Pees, M., Hinojosa, J. C., Lam, N. T., Serracanta, M., Talavera, G., Dincă, V., & Vila, R. (2022). The atlas of mitochondrial genetic diversity for Western Palaearctic butterflies. *Global Ecology and Biogeography*, 31(11), 2184–2190. https://doi.org/10.1111/geb.13579

Dasmahapatra, K. K., Elias, M., Hill, R. I., Hoffman, J. I., & Mallet, J. (2010). Mitochondrial DNA barcoding detects some species that are real, and some that are not. *Molecular Ecology Resources*, 10(2), 264–273. https://doi.org/10.1111/j.1755-0998.2009.02763.x

D'Ercole, J., Dincă, V., Opler, P. A., Kondla, N., Schmidt, C., Phillips, J. D., Robbins, R., Burns, J. M., Miller, S. E., Grishin, N., Zakharov, E. V., DeWaard, J. R., Ratnasingham, S., & Hebert, P. D. N. (2021). A DNA barcode library for the butterflies of North America. *PeerJ*, 9, e11157. https://doi.org/10.7717/peerj.11157

deWaard, J. R., Ivanova, N. V., Hajibabaei, M., & Hebert, P. D. N. (2008). Assembling DNA barcodes. Analytical protocols. *Methods in Molecular Biology (Clifton, N.J.)*, 410, 275–293. https://doi.org/10.1007/978-1-59745-548-0_15

Dincă, V., Backström, N., Dapporto, L., Friberg, M., García-Barros, E., Hebert, P. D. N., Hernández-Roldán, J., Hornett, E., Lukhtanov, V., Marec, F., Montagud, S., Munguira, M. L., Olofsson, M., Šíchová, J., Talavera, G., Vicente-Arranz, J. C., Vila, R., & Wiklund, C. (2015). DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. *Scientific Reports*, 58(8), 391. https://doi.org/10.1139/gen-2015-0123

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. *Communications Biology*, 4(1), 1–11. https://doi.org/10.1038/s42003-021-01834-7

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29(8), 1969–1973. https://doi.org/10.1093/molbev/mss075

Dudley, R., & Srygley, R. (1994). Flight physiology of neotropical butterflies: Allometry of airspeeds during natural free flight. *Journal of Experimental Biology*, 191(1), 125–139. https://doi.org/10.1242/jeb.191.1.125

Ebdon, S., Laetsch, D. R., Dapporto, L., Hayward, A., Ritchie, M. G., Dincă, V., Vila, R., & Lohse, K. (2021). The Pleistocene species pump past its prime: Evidence from European butterfly sister species. *Molecular Ecology*, 30(14), 3575–3589. https://doi.org/10.1111/mec.15981

Edwards, E. (2010). *Studying butterflies and moths of the Samoan archipelago: Manual leaflet and butterfly catalogue* (p. 20). MNRE, JICA.

Fišer, C., Robinson, C. T., & Malard, F. (2018). Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology*, 27(3), 613–635. https://doi.org/10.1111/mec.14486

Fujisawa, T., & Barraclough, T. G. (2013). Delimiting species using single-locus data and the generalized mixed yule coalescent approach: A revised method and evaluation on simulated data sets. *Systematic Biology*, 62(5), 707–724. https://doi.org/10.1093/sysbio/syt033

Galimberti, A., Assandri, G., Maggioni, D., Ramazzotti, F., Baroni, D., Bazzi, G., Chiandetti, I., Corso, A., Ferri, V., Galuppi, M., Ilahiane, L., La Porta, G., Laddaga, L., Landi, F., Mastropasqua, F., Ramellini, S., Santinelli, R., Soldato, G., Surdo, S., & Casiraghi, M. (2021). Italian odonates in the Pandora's box: A comprehensive DNA barcoding inventory shows taxonomic

warnings at the Holarctic scale. *Molecular Ecology Resources*, *21*(1), 183–200. https://doi.org/10.1111/1755-0998.13235

García-Berro, A., Talla, V., Vila, R., Kar Wai, H., Shipilina, D., Chan, K. G., Pierce, N. E., Backström, N., & Talavera, G. (2022). Migratory behavior is positively associated with genetic diversity in butterflies. *Molecular Ecology*, *32*(3), 560–574. https://doi.org/10.1111/mec.16770

Gaunet, A., Dincă, V., Dapporto, L., Montagud, S., Vodă, R., Schär, S., Badiane, A., Font, E., & Vila, R. (2019). Two consecutive *Wolbachia*-mediated mitochondrial introgressions obscure taxonomy in Palearctic swallowtail butterflies (Lepidoptera, Papilionidae). *Zoologica Scripta*, *48*(4), 507–519. https://doi.org/10.1111/zsc.12355

Gillespie, J. H. (2001). Is the population size of a species relevant to its evolution? *Evolution*, *55*(11), 2161–2169. https://doi.org/10.1111/j.0014-3820.2001.tb00732.x

Habel, J. C., Seibold, S., Ulrich, W., & Schmitt, T. (2018). Seasonality overrides differences in butterfly species composition between natural and anthropogenic forest habitats. *Animal Conservation*, *21*(5), 405–413. https://doi.org/10.1111/acv.12408

Hausmann, A., Haszprunar, G., Segerer, A. H., Speidel, W., Behounek, G., & Hebert, P. D. N. (2011). Now DNA-barcoded: The butterflies and larger moths of Germany. *Spixiana*, *34*(1), 47–58.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218

Hebert, P. D. N., Ratnasingham, S., & DeWaard, J. R. (2003). Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences*, *270*(SUPPL. 1), S96–S99. https://doi.org/10.1098/rsbl.2003.0025

Holt, B. J., Lessard, J. P., Borregaard, M. K., Fritz, S. A., Araújo, M. B., Dimitrov, D., Fabre, P. H., Graham, C. H., Graves, G. R., Jønsson, K. A., Nogués-Bravo, D., Wang, Z., Whittaker, R. J., Fjeldså, J., & Rahbek, C. (2013). An update of Wallace's zoogeographic regions of the world. *Science*, *339*(6115), 74–78. https://doi.org/10.1126/science.1228282

Leigh, D. M., van Rees, C. B., Millette, K. L., Breed, M. F., Schmidt, C., Bertola, L. D., Hand, B. K., Hunter, M. E., Jensen, E. L., Kershaw, F., Liggins, L., Luikart, G., Manel, S., Mergeay, J., Miller, J. M., Segelbacher, G., Hoban, S., & Paz-Vinas, I. (2021). Opportunities and challenges of macrogenetic studies. *Nature Reviews Genetics*, *22*(12), 791–807. https://doi.org/10.1038/s41576-021-00394-0

Lescak, E. A., Bassham, S. L., Catchen, J., Gelmond, O., Sherbick, M. L., Van Hippel, F. A., & Cresko, W. A. (2015). Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(52), E7204–E7212. https://doi.org/10.1073/pnas.1512020112

Litman, J., Chittaro, Y., Birrer, S., Praz, C., Wermeille, E., Fluri, M., Stalling, T., Schmid, S., Wyler, S., & Gonseth, Y. (2018). A DNA barcode reference library for Swiss butterflies and forester moths as a tool for species identification, systematics and conservation. *PLoS One*, *13*(12), e0208639. https://doi.org/10.1371/journal.pone.0208639

MacArthur, R. H., & Wilson, E. O. (1967). *The theory of Island biogeography*. Princeton University Press.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, *20*(5), 229–237. https://doi.org/10.1016/j.tree.2005.02.010

Matos-Maraví, P., Núñez Águila, R., Peña, C., Miller, J. Y., Sourakov, A., & Wahlberg, N. (2014). Causes of endemic radiation in the Caribbean: Evidence from the historical biogeography and diversification of the butterfly genus *Calisto* (Nymphalidae: Satyrinae: Satyrini). *BMC Evolutionary Biology*, *14*(1), 199. https://doi.org/10.1186/s12862-014-0199-7

Menchetti, M., Talavera, G., Cini, A., Salvati, V., Dincă, V., Platania, L., Bonelli, S., Balletto, E., Vila, R., & Dapporto, L. (2021). Two ways to be endemic. Alps and Apennines are different functional refugia during climatic cycles. *Molecular Ecology*, *30*(5), 1297–1310. https://doi.org/10.1111/mec.15795

Múrias Dos Santos, A., Cabezas, M. P., Tavares, A. I., Xavier, R., & Branco, M. (2016). TcsBU: A tool to extend TCS network layout and visualization. *Bioinformatics*, *32*(4), 627–628. https://doi.org/10.1093/bioinformatics/btv636

Myers, N., Mittermeier, R., Mittermeier, C., Da Fonseca, G., & Kent, J. (2000). Conservation: Biodiversity as a bonus prize. *Nature*, *403*(7326), 853–858. https://doi.org/10.1038/468895a

Núñez, R., Genaro, J. A., Pérez-Asso, A., Murillo-Ramos, L., Janzen, D. H., Hallwachs, W., Wahlberg, N., & Hausmann, A. (2020). Species delimitation and evolutionary relationships among Phoebis New World Sulphur butterflies (Lepidoptera, Pieridae, Coliadinae). *Systematic Entomology*, *45*(2), 481–492. https://doi.org/10.1111/syen.12408

Patrick, B., & Patrick, H. (2012). *Butterflies of the South Pacific*. University of Otago Press.

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*(3), 554–570. https://doi.org/10.1111/j.0014-3820.2004.tb01678.x

Reynolds, L. A., Hornett, E. A., Jiggins, C. D., & Hurst, G. D. D. (2019). Suppression of *Wolbachia*-mediated male-killing in the butterfly *Hypolimnas bolina* involves a single genomic region. *PeerJ*, *10*, 1–10. https://doi.org/10.7717/peerj.7677

Ritter, S., Michalski, S. G., Settele, J., Wiemers, M., Fric, Z. F., Sielezniew, M., Šašić, M., Rozier, Y., & Durka, W. (2013). *Wolbachia* infections mimic cryptic speciation in two parasitic butterfly species, *Phengaris teleius* and *P. nausithous* (Lepidoptera: Lycaenidae). *PLoS One*, *8*(11), 78107. https://doi.org/10.1371/journal.pone.0078107

Sahoo, R. K., Lohman, D. J., Walhberg, N., Muller, C. J., Brattstrom, O., Collins, S. C., Peggie, D., Aduse-Poku, K., & Kodandaramaiah, U. (2018). Evolution of *Hypolimnas* butterflies (Nymphalidae): Out-of-Africa origin and *Wolbachia*-mediated introgression. *Molecular Phylogenetics and Evolution*, *123*, 50–58. https://doi.org/10.1016/j.ympev.2018.02.001

Salces-Castellano, A., Andújar, C., López, H., Pérez-Delgado, A. J., Arribas, P., & Emerson, B. C. (2021). Flightlessness in insects enhances diversification and determines assemblage structure across whole communities. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1945), 20202646. https://doi.org/10.1098/rspb.2020.2646

Salces-Castellano, A., Patiño, J., Alvarez, N., Andújar, C., Arribas, P., Braojos-Ruiz, J. J., del Arco-Aguilar, M., García-Olivares, V., Karger, D. N., López, H., Manolopoulou, I., Oromí, P.,

Pérez-Delgado, A. J., Peterman, W. E., Rijsdijk, K. F., & Emerson, B. C. (2020). Climate drives community-wide divergence within species over a limited spatial scale: Evidence from an oceanic Island. *Ecology Letters*, *23*(2), 305–315. https://doi.org/10.1111/ele.13433

Scalercio, S., Cini, A., Menchetti, M., Vodă, R., Bonelli, S., Bordoni, A., Casacci, L. P., 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*(9), 2013–2026. https://doi.org/10.1111/1365-2656.13196

Sekar, S. (2012). A meta-analysis of the traits affecting dispersal ability in butterflies: Can wingspan be used as a proxy? *Journal of Animal Ecology*, *81*(1), 174–184. https://doi.org/10.1111/j.1365-2656.2011.01909.x

Shaw, K. L. (2002). Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(25), 16122–16127. https://doi.org/10.1073/pnas.242585899

Shaw, K. L., & Gillespie, R. G. (2016). Comparative phylogeography of oceanic archipelagos: Hotspots for inferences of evolutionary process. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), 7986–7993. https://doi.org/10.1073/pnas.1601078113

Sourakov, A., & Zakharov, E. V. (2011). "Darwin's butterflies"? DNA barcoding and the radiation of the endemic Caribbean butterfly genus *Calisto* (Lepidoptera, Nymphalidae, Satyrinae). *Comparative Cytogenetics*, *5*(3), 191–210. https://doi.org/10.3897/compcytogen.v5i3.1730

Strak, V., & Schellart, W. P. (2018). A subduction and mantle plume origin for Samoan volcanism. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-018-28267-3

Struck, T. H., Feder, J. L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V. I., Kistenich, S., Larsson, K. H., Liow, L. H., Nowak, M. D., Stedje, B., Bachmann, L., & Dimitrov, D. (2018). Finding evolutionary processes hidden in cryptic species. *Trends in Ecology and Evolution*, *33*(3), 153–163. https://doi.org/10.1016/j.tree.2017.11.007

Sucháčková Bartoňová, A., Konvička, M., Marešová, J., Wiemers, M., Ignatev, N., Wahlberg, N., Schmitt, T., & Faltýnek Fric, Z. (2021). *Wolbachia* affects mitochondrial population structure in two systems of closely related Palaearctic blue butterflies. *Scientific Reports*, *11*(1), 1–14. https://doi.org/10.1038/s41598-021-82433-8

Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(7), 1607–1611. https://doi.org/10.1073/pnas.1607921114

Talavera, G., Dincă, V., & Vila, R. (2013). Factors affecting species delimitations with the GMYC model: Insights from a butterfly survey. *Methods in Ecology and Evolution*, *4*(12), 1101–1110. https://doi.org/10.1111/2041-210X.12107

Tennent, W. J. (2006). A checklist of the butterflies of Melanesia, Micronesia, Polynesia and some adjacent areas. *Zootaxa*, *1178*, 1–209. https://doi.org/10.11646/zootaxa.1178.1.1

Theodoridis, S., Rahbek, C., & Nogues-Bravo, D. (2021). Exposure of mammal genetic diversity to mid-21st century global change. *Ecography*, *44*(6), 817–831. https://doi.org/10.1111/ecog.05588

Thomson, L. A. J., & Evans, B. (2006). *Terminalia catappa* (tropical almond). *Species Profiles for Pacific Island Agroforestry*, *2*(2), 1–20.

Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, *21*(16), 3907–3930. https://doi.org/10.1111/j.1365-294X.2012.05664.x

Vellend, M., & Orrock, J. L. (2010). Ecological and genetic models of diversity. In J. B. Losos & R. E. Ricklefs (Eds.), *The theory of Island biogeography revisited* (pp. 439–462). Princeton University Press. https://doi.org/10.1515/9781400831920.439

Vodă, R., Dapporto, L., Dincă, V., Shreeve, T. G., Khaldi, M., Barech, G., Rebbas, K., Sammut, P., Scalercio, S., Hebert, P. D. N., & Vila, R. (2016). Historical and contemporary factors generate unique butterfly communities on islands. *Scientific Reports*, *6*(1), 28828. https://doi.org/10.1038/srep28828

Vodă, R., Dapporto, L., Dincă, V., & Vila, R. (2015). Cryptic matters: Overlooked species generate most butterfly beta-diversity. *Ecography*, *38*(4), 405–409. https://doi.org/10.1111/ecog.00762

Warren, B. H., Simberloff, D., Ricklefs, R. E., Aguilée, R., Condamine, F. L., Gravel, D., Morlon, H., Mouquet, N., Rosindell, J., Casquet, J., Conti, E., Cornuault, J., Fernández-Palacios, J. M., Hengl, T., Norder, S. J., Rijsdijk, K. F., Sanmartín, I., Strasberg, D., Triantis, K. A., … Thébaud, C. (2015). Islands as model systems in ecology and evolution: Prospects fifty years after MacArthur-Wilson. *Ecology Letters*, *18*(2), 200–217. https://doi.org/10.1111/ele.12398

Whittaker, R. J., & Fernández-Palacios, J. M. (2007). *Island biogeography: Ecology, evolution, and conservation*. Oxford University Press.

Whittaker, R. J., Fernández-Palacios, J. M., Matthews, T. J., Borregaard, M. K., & Triantis, K. A. (2017). Island biogeography: Taking the long view of natures laboratories. *Science*, *357*(6354), eaam8326. https://doi.org/10.1126/science.aam8326

Wiemers, M., Balletto, E., Dinca, V., Fric, Z. F., Lamas, G., Lukhtanov, V., Munguira, M. L., VanSwaay, C. A. M., Vila, R., Vliegenthart, A., Wahlberg, N., & Verovnik, R. (2018). An updated checklist of the European butterflies (Lepidoptera, Papilionoidea). *ZooKeys*, *811*, 9–45. https://doi.org/10.3897/zookeys.811.28712

Zamani, A., Fric, Z. F., Gante, H. F., Hopkins, T., Orfinger, A. B., Scherz, M. D., Bartoňová, A. S., & Pos, D. D. (2022). DNA barcodes on their own are not enough to describe a species. *Systematic Entomology*, *47*(3), 385–389. https://doi.org/10.1111/syen.12538

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.