Genomics, population divergence and historical demography of the world's largest and endangered butterfly, the Queen Alexandra's birdwing

Eliette L Reboud, Benoit Nabholz, Emmanuelle Chevalier, Marie-Ka Tilak, Darren Bito, Fabien L Condamine

Research Article

Title

Genomics, population divergence and historical demography of the world's largest and endangered butterfly, the Queen Alexandra's birdwing

Authors

Eliette L. Reboud[1]*, Benoit Nabholz[1,2], Emmanuelle Chevalier[1], Marie-ka Tilak[1], Darren Bito[3] & Fabien L. Condamine[1]*

Affiliations

[1] Institut des Sciences de l'Evolution de Montpellier (Université Montpellier | CNRS | IRD | EPHE), Place Eugène Bataillon, 34095 Montpellier, France

[2] Institut Universitaire de France (IUF), Paris, France

[3] Pacific Adventist University, Private Mail Bag, BOROKO 111, National Capital District, Papua New Guinea

Corresponding authors (*)

Eliette L. Reboud: eliette.reboud@gmail.com

Fabien L. Condamine: fabien.condamine@gmail.com

Running head

Genomics of the world's largest butterfly

Abstract

The world's largest butterfly is the microendemic Papua New Guinean *Ornithoptera alexandrae*. Despite years of conservation efforts to protect its habitat and breed this up-to-28-cm butterfly, this species still figures as endangered in the IUCN Red List and is only known from two allopatric populations occupying a total of only ~140 km². Here we aim at assembling reference genomes for this species to investigate its genomic diversity, historical demography and determining whether the population is structured, which could provide guidance for conservation programs attempting to (inter)breed the two populations. Using a combination of long and short DNA reads and RNA sequencing, we assembled six reference genomes of the tribe Troidini, with four annotated genomes of *O. alexandrae* and two genomes of related species *O. priamus* and *Troides oblongomaculatus*. We estimated the genomic diversity of the three species, and we proposed scenarios for the historical population demography using two polymorphism-based methods taking into account the characteristics of low-polymorphic invertebrates. Indeed, chromosome-scale assemblies reveal very low levels of nuclear heterozygosity across Troidini, which appears to be exceptionally low for *O. alexandrae* (lower than 0.01%). Demographic analyses demonstrate low and steadily declining *Ne* throughout *O. alexandrae* history, with a divergence into two distinct populations about 10,000 years ago. These results suggest that *O. alexandrae* distribution has been microendemic for a long time. It should also make local conservation programs aware of the genomic divergence of the two populations, which should not be ignored if any attempt is made to cross the two populations.

Key words:

Conservation genomics, heterozygosity, low genetic diversity, *Ornithoptera alexandrae*, reference genome.

Significance statement

Despite its charisma, little is known about the demographic trends and taxonomic status of the two populations from the giant endangered birdwing butterfly *Ornithoptera alexandrae*. By sampling and sequencing individuals of this species and two closely related species, we study whether and how the population is structured, and we investigate the genomic diversity of the species and the "health" of their genomes and populations (e.g., demographic trend, evidence of inbreeding). Overall, the very low genomic diversity and steadily declining trend inferred by this study suggests that efforts need to be reinforced to conserve this amazing Papua New Guinean insect.

Introduction

When in January 1906, Alfred S. Meek saw an enormous butterfly flying high above him in the canopy of this forest of the Northern Province of Papua New Guinea, two days walk from the coast, he took his rifle and shot down the beast. This is what one can read in the letter he sent to his correspondent Karl Jordan at the Natural History Museum of Tring (England) (letter n°155 of Meek's communications, Meek 1906; Ackery 1997; Tennent 2021). Meek let his funder, Lord Walter Rothschild describe in 1907 for the first time *Ornithoptera alexandrae* (Papilionidae: Troidini), known as the Queen Alexandra's birdwing butterfly, based on this female whose wing still bears the stigma of this extraordinary hunt. This “trophy” and the even larger congeners that followed have become representatives of the world's largest known butterfly species to date and contribute to the continuing amazement of scientists at the incredible diversity, size, and beauty of Papua New Guinea’s insects (Parsons 1992; Mitchell et al. 2016). Indeed, many naturalists have been studying *O. alexandrae*, culminating with the comprehensive review on this butterfly by Mitchell et al. (2016) that serves as a basis for this work.

As the world’s largest butterfly, *Ornithoptera alexandrae* can measure up to 28-30 cm in wingspan (Mitchell et al. 2016 and references therein). *Ornithoptera alexandrae* is endemic to the Northern Province of Papua New Guinea, in a narrow range around Popondetta (Northern Province; **Fig. 1**). Long-term field observations in the last decades have shown there are two recognized allopatric populations: a lowland population in Popondetta plains (≤300 m above sea level), and a highland population occurring on the relatively inaccessible Managalas Plateau about 800 m above sea level (Collins and Morris 1985; Parsons 1999; Böhm 2018). A mountain range separates the two populations, bounded in the West by Mount Lamington volcano (1700 m), and eastward to a mountain 2,140 m high. According to available data, there have been no sightings in between. The volcanic activity of Mount Lamington (last eruption in

1951 with activity until 1956; Global Volcanism Program, 2022), the flooding, drought and fires occurring in the region as well as recent logging and agricultural activities might explain today's fragmented distribution of *O. alexandrae* (Parsons 1992; Mitchell et al. 2016). The relatively small distribution range composed of two patches has been interpreted by Haugum and Low (1979) as a relict occurrence, potentially due to an evolutionary bottleneck or demographic decline. However, genetic studies are crucially lacking to assess this hypothesis.

Ornithoptera alexandrae is considered as a threatened species in the IUCN Red List (Böhm 2018). While this species is very rare over an area of occurrence of 8,710 km², its actual area of occupancy is only 128-140 km² fragmented in two populations, placing it in the Endangered category (Böhm 2018). There are doubts about the previously thought monophagy of its caterpillar due to the lack of comprehensive morphological or genetic studies on the Southeast Asian genus *Aristolochia* (Parsons 1996; Buchwalder et al. 2014; Mitchell et al. 2016). It has long been thought that there were no particular restriction due to host plant distribution to explain the local occurrence of *O. alexandrae* as there are many areas where the main larval food plants (previously thought to be *Aristolochia dielsiana*, possibly being *A. meridionaliana* in the highlands and *A. alexandriana* in the lowlands) grow prolifically. This pattern has been described in some monophagous lepidopteran species (Quinn et al. 1997) and may highlight the fact that the distribution of *O. alexandrae* species is also driven by other factors of a microclimatic, pedologic or geological nature that might limit its distribution. On the other hand, some factors represent a threat or vulnerability for the species: it is a large species with a relatively specialized ecology (larvae on a single or very few host-plants and adults in a single habitat) (Koh et al. 2004, Palash et al. 2022), it is sympatric to other *Ornithoptera* species (Koh et al. 2004), and its habitat is or has been fragmented by fires, droughts, and volcanic eruptions, and is severely affected by agriculture (Parsons 1992; Mitchell et al. 2016) and other human activities, with habitat conversion leading to a local

decline in larval host vines. Accordingly, *O. alexandrae* is placed at the top of the CITES list (Appendix I). However, although its international trade is prohibited, this species is highly prized and is subject to an illegal trade that is dangerous for the demography of the species, and its survival (Mitchell et al. 2016). Thus, better monitoring of this species is recommended by the IUCN to better track population trajectories (Böhm 2018), particularly because its numbers and current trend in population dynamics are unknown.

The two allopatric populations of *O. alexandrae* are externally morphologically similar but express important biological differences, such as slight differences in size (on average 14% larger in highlands) and development time (on average 34.5% longer in highlands) (Straatman 1971; Mitchell et al. 2016). The soil and host plant eaten by the larvae might also differ in the two populations (Haugum and Low 1979; Mitchell et al. 2016), so it is unclear how divergent these two populations are. In fact, genomic research on other butterfly groups has revealed that superficial similarity in adults can hide a previously unrecognized cryptic lineage (e.g., Hebert et al. 2004; Burns et al. 2008). Knowledge of evolutionary and genetic history of the species and populations could help conservation and breeding programs to save the species. Genomics is considered as a powerful tool for studying the past and present structure and diversity of populations and brings an invaluable source of information, especially of species that are naturally rare and difficult to study (e.g., Westbury et al. 2018; Van der Valk et al. 2020; Morin et al. 2021). Genome sequencing is increasingly recognized as an important contribution to identifying, characterizing, and conserving biodiversity (Formenti et al. 2022). Reference genomes provide primary data for understanding historical demography (Morin et al. 2021), gene and trait evolution (Warren et al. 2021), or even susceptibility to inbreeding depression and accumulation deleterious mutations (Chattopadhyay et al. 2019; Van der Valk et al. 2020; Robinson et al. 2022). Genomic resources are also useful for broader studies of population structure, relatedness, and recovery potential (e.g., Garner et al. 2016; Morin et al. 2018;

Tunstall et al. 2018), or for assessing correlations between current IUCN status and past demography (Nadachowska-Brzyska et al. 2015). These types of estimates (e.g. sequentially Markovian coalescent [SMC] methods, Li and Durbin 2011) have been widely used for conservation purposes for vertebrates, such as mammals (Morin et al. 2021) or birds (Nadachowska-Brzyska et al. 2015) and insect pests (Hazzouri et al. 2020; You et al. 2020). Despite the continuous increase of threatened insects (Sánchez-Bayo and Wyckhuys 2019), it has been much less used for insect conservation (but see Mikheyev et al. 2017; Podsiadlowski et al. 2021). It is indeed challenging to study the demography of invertebrates using polymorphism-based methods because the risk of violating the assumptions of SMC-type models is high. For instance, Sellinger et al. (2021) revealed that these methods of inference perform poorly when the ratio between the recombination and mutation rates is important, therefore highlighting that the consideration of this ratio is crucial and still much too little considered in this type of analyses in the literature.

Here we perform a first genomic study of *O. alexandrae* to understand the past and present demography of this species and to bring insights into its Endangered status, which may have implications for conservation strategies. Since a local conservation program has been set up and ongoing to rear the two populations, the taxonomic status of these two populations (i.e., populations or species) may inform conservation management of this threatened species (Mitchell et al. 2016). If the two populations are too divergent, it could be complicated to breed specimens from the Managalas Plateau with specimens from the Popondetta lowlands. Given the above-mentioned biological features of this butterfly species, we aim to: (1) assemble high-quality and annotated whole genomes for the two populations, (2) assess the level of nuclear heterozygosity, (3) estimate the demographic history of the species and the two populations in relation to past environmental change and to test whether the current range of the species is

relictual, and (4) provide information for the policy makers to improve their conservation strategy.

Results and Discussion

High-quality assemblies for birdwing butterflies

We collected live specimens from the two populations of *Ornithoptera alexandrae* with one adult and one caterpillar per population and sequenced the DNA combining a mean of 72.5x of long reads (Oxford Nanopore) for draft assembly, 82x of short reads (Illumina) for polishing, as well as 51.7 million cleaned RNAseq reads (8.4 Gb) for genome annotation (see *Materials and Methods*). Using Flye assembler (Kolmogorov et al. 2019) and Pilon polisher (Walker et al. 2014), we assembled the four genomes of *O. alexandrae* that range from 321 to 327 Mb, which are very contiguous with a mean of 582 contigs (ranging from 305 to 1222 contigs) and a mean N50 of 9.9 Mb (**Table 1**). Over a total of 5,286 core genes of the Lepidoptera database (odb10, Manni et al. 2021), BUSCO recovered on average 98.83% ±0.05 complete genes, 0.23% ±0.05 fragmented genes and 0.97% ±0.06 missing genes (**Table 1**). The genome size and gene completeness of our *O. alexandrae* assemblies are comparable to the genome of a related species: *Troides helena* (330 Mb, BUSCO score = 96.6%), which was assembled with similar data and methods (He et al. 2022). Furthermore, the genome size stands among the smallest within the family Papilionidae but still is 30 to 40% larger than some *Papilio* (the sister tribe of Troidini; He et al. 2022; Liu et al. 2020) illustrating the dynamic genome size evolution of the family.

After removing potential exogenous contigs of the assemblies (see *Materials and Methods*), we selected FC563 as the reference genome for further analyses, as it had the best assembly statistics and we found no evidence of contamination (i.e., bacteria, plants). We assessed the quality of this *O. alexandrae* FC563 assembly by looking at its correspondence

with the reference genome of *Papilio bianor* (chromosome-level assembly, Lu et al. 2019). We found 24 contigs that match with more than 70% of the length of *P. bianor* chromosomes out of 30 chromosomes. This represents a cumulative length of 81% of the total assembly. Among those, 18 contigs have a single reciprocal match with one *P. bianor* chromosome (representing a cumulative length of 48% of the assembly). One of the most fragmented chromosomes is the Z chromosome in which 11 contigs of *O. alexandrae* FC563 assembly match to chromosome 30 (Z) of *P. bianor*. This is not surprising as FC563 is a female and, therefore, has half the coverage on the Z compared to the autosome. However, a similar analysis performed on the male FC560 led to eight contigs matching to chromosome 30 (Z) of *P. bianor*, suggesting that chromosome Z is difficult to assemble. These eight contigs linked to the Z chromosome were independently identified using coverage and heterozygosity information in FC563, representing a cumulative length of 14.1 Mb (supplementary table S1, Supplementary Material online). The FC563 assembly is therefore composed of 24 full-length or nearly full-length chromosomes, including 18 full-length chromosomes (supplementary figure S1, Supplementary Material online). Analysis of the chromosome-level synteny between *O. alexandrae* and *P. bianor* shows a high level of genomic synteny (supplementary figure S1, Supplementary Material online). Our results suggest that the combination of Nanopore long reads and Illumina short reads perform notably well to recover chromosome-scale assemblies as we reconstructed genome assemblies comparable to the best assemblies of Papilionidae available so far (comparison with *Papilio bianor*, Lu et al. 2019: table 1) without relying on Hi-C techniques.

Using transcriptomic data, protein homology and *de novo* genes prediction, we annotated the genomes of the two populations (FC560 and FC653) using the MAKER pipeline (Holt and Yandell 2011). Protein predictions retrieved 17,617 genes for FC560 and 16,508 genes for FC563. We carried out BUSCO analyses with these two proteoms, and estimated

97.7% complete genes, 0.9% fragmented genes and 1.4% missing genes for FC560, and 97.2% complete genes, 1.0% fragmented genes and 1.8% missing genes for FC563. Because FC560 annotation contained more genes, it was transferred to the other two genomes of *O. alexandrae*. Overall, 33-35% of the genome was annotated as repeat sequences with mostly unclassified categories of interspersed repeats (supplementary table S2, Supplementary Material online). This proportion of repeats is relatively high compared to other Papilionidae genomes already available (22% in *Papilio glaucus*, 22.4% in *P. xuthus*; Cong et al. 2015; Lu et al. 2019), except for the species with larger genomes such as *Papilio bianor* (55%; Lu et al. 2019) and *Parnassius apollo* (65%; Podsiadlowski et al. 2021). This is consistent with a positive correlation between assembly size and repeat content in insects (Petersen et al. 2019; Heckenhauer et al. 2022; Sproul et al. 2022). Within *O. alexandrae*, all individuals have similar cumulative repeat size, and no difference was detected between lowland and highland individuals despite the fact that highland individuals had a slightly larger genome assembly than lowland individuals (~5 Mb, supplementary table S2, Supplementary Material online). These annotated genomes are available in GenBank (Bioproject PRJNA938052).

For comparison purposes, we also sequenced and assembled two other birdwing butterfly species (Troidini): *Ornithoptera priamus poseidon* and *Troides oblongomaculatus papuensis*, both ranked as Least Concern in the IUCN Red List (see *Materials and Methods*). Both genomes were very similar to the *O. alexandrae*'s ones in quality and assembly statistics (**Table 1**).

The world's largest butterfly has low levels of genomic diversity

Using short reads data and annotation with MitoFinder (Allio et al. 2020) (see *Materials and Methods*), each mitogenome was reconstructed in a single contig and we retrieved 36-37 genes (including 13 protein-coding genes and 2 rRNA), with evidence of circularization. The

mitogenomic diversity (π-diversity) including coding and non-coding region (such as the D-loop) was calculated at ~0.0704% (supplementary table S3, Supplementary Material online). This is comparable to the low level of mitochondrial diversity for mammalian species such as the Tasmanian devil or Island fox (Westbury et al. 2018). Up to our knowledge, there is no estimation of mitochondrial diversity based on whole mitogenomes of butterflies, but there are studies estimating mitochondrial diversity relying on the cytochrome *c* oxidase subunit I (COI) DNA barcode marker (e.g., π-diversity: Mackintosh et al. 2019; haplotype diversity: Dincă et al. 2021). The mitochondrial diversity of the DNA barcode for *O. alexandrae* is ~0.076% (supplementary table S3, Supplementary Material online), which ranks fifth lowest of 38 European butterflies (Mackintosh et al. 2019). However, a low mitochondrial diversity does not equate low autosomal diversity (Allio et al. 2017; Mackintosh et al. 2019).

Using mapping data on the best genome assembly of *O. alexandrae* FC563, we calculated the autosomal heterozygosity of the two populations. All four individuals of *O. alexandrae* show a heterozygosity around 0.08% (supplementary table S4, Supplementary Material online). Using a similar metric relying on four-fold degenerate sites from coding sequences, this heterozygosity is the lowest value observed within the butterfly studied by Mackintosh et al. (2019) (**Fig. 2,** supplementary table S4, Supplementary Material online). We also performed sensitivity analyses to corroborate these results using different data types (short reads vs. long reads) and window sizes (100 kb vs. 1 Mb) and proportion of missing data (10% vs. 20%). Overall, the heterozygosity of *O. alexandrae* ranges from 0.0745 to 0.0838% (supplementary table S4, Supplementary Material online), suggesting the low estimated heterozygosity is not an artifact due to methods and/or data. Determinants of heterozygosity are still a long-term debate (Romiguier et al. 2014; Galtier and Ellegren 2016; Mackintosh et al. 2019; Buffalo 2021) and intrinsic features of *O. alexandrae* like its large size, its tropicality, or its microendemism might explain such a low heterozygosity. However, for two closely

related species having similar lifestyle but larger distributional range and abundance than *O. alexandrae* (**Fig. 3)**, we found that *O. priamus* has a higher heterozygosity rate (autosomal 0.433%, neutral diversity 0.708%, ca. 6 and 10 times higher) and that *T. oblongomaculatus* has an even lower heterozygosity than *O. alexandrae* (0.0737% autosomal, 0.0704% neutral). Interestingly, these levels of heterozygosity are much lower than the level estimated in the genus *Papilio* ranging from 1.0 to 2.3% although estimated with a different method (k-mer distribution; Lu et al. 2019). Accordingly, the low level of heterozygosity cannot be explained by the species' range or body size, as suggested in Mackintosh et al. (2019) (**Fig. 3**).

Historical demography

Although the causes of the low genetic diversity of these birdwing butterflies are unclear, we can wonder how this translates into demographic history of the species given the threats they are experiencing. Predictions can be formulated such as a prolonged decline of effective population size such as inferred for the brown hyena and the Californian condor (Westbury et al. 2018; Robinson et al. 2021) or a low but stable effective population size such as in the vaquita porpoise (Morin et al. 2021; Robinson et al. 2022). To our knowledge, there are still few examples of demographic history in insects, other than pest insects (e.g., red-palm weevil, Hazzouri et al. 2020; diamondback moth, You et al. 2020; but see Walton et al. 2021; Manthey et al. 2022; García-Berro et al. 2023). Within swallowtail butterflies, the recent study of the Apollo butterfly (*Parnassius apollo*), ranked as a threatened species, showed using SNP data synchronous population declines throughout different mountain massifs despite high heterozygosity levels (Kebaïli et al. 2022).

Sequential Markovian coalescent models (e.g. PSMC, MSMC) are widely used to study the trajectory of the ancestral effective population size (*Ne*) over time. We applied the MSMC2 model to the best genome assembly of *Ornithoptera alexandrae* as well as *O. priamus* and

Troides oblongomaculatus before focusing on the two populations of *O. alexandrae*. SMC models generally infer a *Ne* under the panmictic assumption. This strong assumption is often false in reality, and theoretical work and simulations have shown that SMC dynamics might also be caused by population structure and connectivity changes (Teixeira et al. 2021), so this should be borne in mind when interpreting these population size analyses (Bentley and Armstrong, 2021; Teixeira et al. 2021). Furthermore, it has been shown that SMC methods do not perform well when the mutation rate μ is rarer than the recombination rate r (Sellinger et al. 2021). Unfortunately, this is likely to be very common in many groups such as fungi, sea cucumbers, nematodes, corals, insects, and plants (see Wilfert et al. 2007; Stapley et al. 2017 for recombination rates and Lynch et al. 2016; Wang & Obbard, 2023 for mutation rates). In practice, there is concern that, at least many small non-vertebrate genomes have a much lower μ than r (Sellinger et al. 2020, Sellinger et al. 2021) whose demographic inferences would be affected and poorly addressed in the literature (e.g. coral: Fuller et al. 2020; insects: Waldvogel et al. 2017; Hazzouri et al. 2020; You et al. 2020; Wang et al. 2022; Garcia-Berro et al. 2023).

Demographic inferences on simulated data with the range of recombination and mutation rate parameters of *O. alexandrae* (i.e., r about ten times higher than μ, see *Materials and Methods*) confirmed the methodological issue, which was almost completely reduced when the model was set with an appropriate initial value of the ratio (**Fig 4**). We do not know exactly how the model behaves with an appropriate initial value on real data, but our results changed significantly when the initial value was adjusted (supplementary figure S2, Supplementary Material online), highlighting that this is probably a very important parameter to consider in SMC analyses and further emphasizing the importance of taking the results of any SMC analyses with caution.

Finally, MSMC2 analyses traced relatively different demographic histories for the three butterflies (**Fig. 5**). The effective population size of *O. alexandrae* seems to have been at a low but continuously declining number from ~250,000 to 50,000 individuals during the last million year. However, we inferred a recent demographic change in the last 10,000 years (see below). This demography dynamic is similar to that of the vaquita porpoise, whose effective population size has never been elevated but remained stable for much of its known history and now has one of the lowest rates of heterozygosity known among marine mammals (Morin et al. 2021; Robinson et al. 2022). Altogether, these results suggest that the ancestral effective population of *O. alexandrae* has never been large, thus suggesting that it has been microendemic for a very long time. Interestingly, *T. oblongomaculatus* experienced a similar fairly low and regular *Ne* during most of its history with a *Ne* that culminated (at ~0.3 million individuals) around 10,000 years ago (**Fig. 5**). It was followed by a decreasing trend toward the present, which seems consistent with the low rate of heterozygosity that we estimated today. In contrast, *O. priamus* shows an early peak of *Ne* (~2.0 million individuals ~250,000 years ago) followed by a severe decline until ~40,000 years ago.

In the literature, temporal variations of *Ne* are usually compared with past climatic fluctuations such as temperature and/or sea level, in particular in line with Quaternary glaciations (e.g., Nadachowska-Brzyska et al. 2015; Westbury et al. 2018; Hazzouri et al. 2020; Morin et al. 2021; Teixeira et al. 2021). Although it is tempting to associate the inferred *Ne* variations of the studied birdwing butterflies with the glaciation-interglaciation cycles, it remains difficult to extract a correlation because of climatic heterogeneity at the global scale and estimates of demographic parameters (Bentley and Armstrong, 2022). The last million years was mostly a glacial period that have also been documented in New Guinea (e.g., Chappell and Polach 1991; Barrows et al. 2011), punctuated by intermittent warming events

(e.g., Eemian event ~115,000-130,000), with the last glacial period (Würm glaciation event) starting ~115,000 and ending ~11,700 years ago (last glacial maximum at 19,000-20,000 years). During the last glacial maximum, temperatures were 5°C colder (Barrows et al. 2011) and sea level was 70 m lower (Chappell and Polach 1991) than today. The last glaciation period coincides with the decrease of *Ne* for *O. priamus*. On the contrary, the scale induced by *O. priamus*' inference does not allow to discern any congruence between climatic changes and the *Ne* of *T. oblongomaculatus* and *O. alexandrae*. In addition, this result relies only on a single genome for *O. alexandrae* whereas the two populations of *O. alexandrae* might have had different demographic histories.

Within *O. alexandrae*, every individual showed a similar demographic pattern (low and steadily declining *Ne* during most of the history, **Fig. 6a**). Nonetheless, the two populations of *O. alexandrae* seem to separate initially ~10,000 years ago, and have likely been followed by a relative increase of the *Ne* in the last period (after the split). These results were validated when using multiple genomes per population (see *Materials and Methods*) (**Fig. 6b**) and computed the cross coalescence rate (CCR) between populations (**Fig. 6c**). The CCR gives the probability that a coalescence happens between rather than within the populations and, therefore, quantifies isolation (Wang et al. 2020). Using a generation time of 0.75 years, the CCR started to decrease ~10,000 years ago (CCR = 0.95) and was below 5% at ~1,000 years ago (CCR = 0.05) (**Fig. 6c**). These results suggest that two populations have recently diverged, potentially just after the last glaciation, raising the question of the extent to which the two populations are fully isolated or whether some gene flow (migration) is still ongoing.

Population structure

To explore the recent divergence between the two populations, we computed the median F_{ST} on sliding windows of 100 kb across the whole genome (see *Materials and Methods*). We

estimated a median F_{ST} of 0.03 but the F_{ST} was highly variable among genomic positions with top 5% windows having F_{ST} above 0.14 (**Fig. 7**). The F_{ST} and nucleotide diversity clearly decrease in the center of the large contigs in regions (**Fig. 7**) as expected along chromosomes (e.g., Tine et al. 2014; Delmore et al. 2018). This large-scale variation in F_{ST} and genetic diversity, repeated among scaffolds, suggest the action of endogenous effects such as recombination rate variation, rather than the effect of exogenous local adaptation (Burri 2017). This is consistent with reduced recombination rate in centers and extremities of the chromosomes as recently reported in the painted lady (*Vanessa cardui*; Shipilina et al. 2022). In addition, the cumulative size of 100-kb windows with less than two heterozygous sites represents less than 1% of the genome in all individuals. Moreover, not a single window of 1 Mb with one heterozygous site or less was detected in any individuals. It suggests that the genome of *O. alexandrae* shows no sign of recent inbreeding.

Using the approximate Bayesian computation (ABC) framework implemented in DILS (Fraïsse et al. 2021), we fitted the four models (secondary contact [SC], strict isolation [SI], isolation with migration [IM], ancient migration [AM]) across randomly selected genomic regions to estimate population sizes, migration rates and time of population separation according to each model (i.e., the date when gene flow stopped) between lowland and highland populations (see *Materials and Methods*). On the four runs, DILS only set aside the secondary contact model, but could not identify a best-fitting model between the three others: (2 runs supported migration (IM model), and 2 runs found isolation (1 run for AM and 1 run for SI). Depending on the best fitted model, the time of split was either more recent than MSMC (isolation) or similar (migration) and varied from 5,100 generations (~3,825 years ago) to 16,850 generations (~12,637 years ago) for the median of the posterior distribution of the SI model and the IM model respectively (Supplementary table S5, Supplementary Material online).

Despite the uncertainty of the DILS scenarios, notably on the presence or intensity of migratory flows between the two populations, all the analyses seem overall to agree on the rather consistent isolation of the two populations for a few thousand years. This is interesting because field observations tend to say that the two populations are externally morphologically similar but express biological differences. The highland population is on average larger than the lowland population (264 vs. 218 mm wingspan on average; Mitchell et al. 2016). Furthermore, *O. alexandrae* populations may feed on two different *Aristolochia* species, which could be *A. meridionaliana* in the highlands and *A. alexandriana* in the lowlands (Haugum and Low 1979; Parsons 1996; Mitchell et al. 2016). However, it is difficult to reach a conclusion on this point since little is known on the taxonomy of these two plant morphs and, as mentioned by Buchwalder et al. (2014) who started a taxonomic revision of the genus Papuan *Aristolochia*, there is still significant work to be done to update the nomenclature of this group. The two populations of *O. alexandrae* also differ slightly in their life cycle. From the egg, through larval and pupal stages to adults, the life cycle of *O. alexandrae* takes about 131 days (74 days as larva) in the lowland, while it takes 200 days (125 days as larva) in the highland of the Managalas Plateau (Straatman 1971; Mitchell et al. 2016). Such a difference may be explained by a temperature gradient between the low and highlands (up to 4°C cooler than the plain).

Overall, our results indicate a lack of evidence of a strong genomic differentiation and a genomic barrier to gene flow between the two allopatric populations of *O. alexandrae* that could justify the description of two species. Nevertheless, given the morpho-ecological differences and the fact that there is a modest population structure (F_{ST}) and divergence time probably initiated after the Last Glacial Maximum, it is possible that the populations deserve subspecies status. However, determining whether the observed morpho-ecological differences are fixed in the genome and to what extent the genitalia differ would shed more light on this

proposal, although the latter is difficult to satisfy due to the lack of available individuals to dissect for this threatened and protected species.

Conservation implications

Conservation today is often approached as a triage, leaving out some "unsalvageable" species in an attempt, at best, to save others (Hayward and Castley 2018). As discussed by Wiedenfeld et al. (2021), this approach to the sixth mass extinction should be primarily a matter of prioritizing the allocation of economic and material resources to the rescue of any species, and the current system is failing to allocate enough resources to save biodiversity. The choice to help certain species rather than others is risky since this choice is necessarily based on partial knowledge and limited data, such as their apparent rarity, sometimes interpreted as an abrupt decline, of certain species. However, it seems that comprehensive and meticulous works of some studies such as the one carried out on the kākāpō (Dussex et al. 2021), Island fox (Robinson et al. 2018) or the vaquita porpoise (Morin et al. 2021, Robinson et al. 2022) in recent years are a further argument that species are not necessarily doomed to extinction due to their rarity and low numbers (e.g., Robinson et al. 2018; Wiedenfeld et al. 2021; Robinson et al. 2022). On the contrary, these rare species may have been rare for so long that this may have led them for instance to have purged more deleterious recessive alleles, they might not suffer from inbreeding depression, and their low number may be less of a problem in recovering their population than many other species that would appear to us less endangered at first sight. Protection measures are proposed and sometimes implemented as a result of scientific studies revealing the urgency of acting to save species. Most of these measures concern vertebrates (e.g., large mammals), often omitting threatened insect diversity from conservation policies (but see Mikheyev et al. 2017). As Harvey et al. (2020) advocates in their roadmap for insect conservation, there is a range of actions to be taken, from short-term actions such as drastically

or immediately stopping threats, to medium- and long-term actions such as conducting new research to fill the knowledge gap regarding insect decline or launching sustainable funding initiatives with the aim of restoring, protecting and creating new vital habitats for insects.

Ornithoptera alexandrae is one of the most charismatic invertebrate species and is receiving special attention as a flagship species for invertebrate conservation (Mitchell et al. 2016). The biological characteristics of *O. alexandrae* correspond to a wide range of traits identified as vulnerability factors, notably its large size, its host and environment specificity, the vulnerability of its host plant or the presence of inter-specific competition in its distribution range (Koh et al. 2004; Mattila et al. 2006). On the other hand, the relatively similar past evolutionary patterns and dynamics of *O. alexandrae* with the vaquita porpoise, showing a low heterozygosity and a low *Ne* for a long time, suggest that we can have hope about the capacity of *O. alexandrae* to avoid extinction. However, the species experienced declining trends in the past and these overall results should not be an argument to relax the protection measures. On the contrary, we take the opportunity to encourage the authorities and associations for increasing protection actions, especially since we know very little about the health of the butterfly populations for the last 4,000 years, while the destruction of its habitat has clearly not decreased since. Logging opens up previously inaccessible areas, resulting in changes in *O. alexandrae* habitat by increasing the number and extent of clearings of primary and secondary forests (Mitchell et al. 2016). Today, oil palm exploitation likely constitutes threats through planned or ongoing deforestation, as well as the increased human population on the Popondetta Plain associated with this activity (Mitchell et al. 2016). Given the difficulty to access the Managalas Plateau, the anthropic pressure is today higher in lowlands, suggesting that the lowland population is experiencing more threats. Therefore, we urge policy makers to (1) protect the rainforest environment of *O. alexandrae*, especially in the highlands where a specific protection area could be exclusively dedicated to it, and (2) strongly support the

lowland population through the ongoing rearing program with breeding and release in the wild. Both proposals are discussed and analyzed in detail in Mitchell et al. (2016) and we fully agree with their statement on the urgency of action to protect *O. alexandrae* and how to address the problem. Note that we advise avoiding interbreeding between the two populations as there is an ongoing process of natural evolution leading to genetic divergence between the two populations.

An interesting research perspective would be to perform a temporal sampling of genomic data from museum specimens. This can provide a more accurate approach to quantify genetic threats in endangered species and to estimate recent decreases in genome-wide diversity, increases in inbreeding levels, and accumulation of deleterious genetic variation (Díez-del-Molino et al. 2018) like it has been done on kākāpō (Dussex et al. 2021) and the vaquita porpoise (Robinson et al. 2022).

Materials and Methods

Sampling, DNA and RNA extractions and sequencing

With permits, a total of four individuals of *Ornithoptera alexandrae* were collected in Papua New Guinea (Oro Province) in November 2019 by F.L.C. and D.B. Two specimens per population were collected: one female and one caterpillar for the lowland population (near Popondetta), and one male and one caterpillar for the highland population in the Managalas Plateau (near Kawowoki village). For each specimen, head, thorax, and abdomen have been separated with scalpels, crushed with surgical scissors, and conserved separately in RNAlater, then stored in freezers at -20°C after a few days at ambient temperature (and one day at 4°C as a transition). One individual of *Ornithoptera priamus* and one individual of *Troides oblongomaculatus* have been collected under similar conditions during the same mission.

Tissues from the thorax or abdomen were used to extract high molecular weight DNA. As part of tests that ended up being similar in terms of quality of sequencing, we used two different extraction methods. The first was the phenol-chloroform method, including a specific ratio of 0.8x AMPure beads applied to retain the longest DNA fragments (Tilak et al. 2020). The second method was the use of the Qiagen genomic DNA kit. This second solution was ultimately applied to most samples due to a better 260/230 ratio in Nanodrop assays, as DNA purity is essential for long-read sequencing, especially for Oxford Nanopore Technology (ONT) sequencing. In addition, one of the four samples of FC561 was treated with the Short Read Eliminator Kit XS (Circulomics, PacBio, USA) to discard sequences below 10 kb long. Final DNA purity and concentrations were measured using Nanodrop (Thermo Fisher, USA) and Qubit (Thermo Fisher, USA). RNA was extracted for *O. alexandrae* only. Extraction and purification were conducted with the Qiagen RNeasy kit. We used part of the thorax of caterpillars (FC561/FC563) and part of the abdomen for adults (FC560/FC562), that were crushed in the lysis buffer.

Library preparations and sequencing

Whole-genome libraries were constructed using the resulting high-molecular-weight DNA as input for the Nanopore LSK-109 ligation kit (Oxford Nanopore Technologies, UK) following the manufacturer's protocol. Long-read sequencing was performed on a GridION device with two to four R9.4.1 flow cells, depending on the individuals (Supplementary table S6, Supplementary Material online).

Remaining DNA extractions of each individual were sent to Novogene Europe (Cambridge, UK) for library preparations. Libraries were generated using NEBNext DNA Library Prep Kit following manufacturer's recommendations and indices were added to each sample. Genomic DNA was randomly fragmented to a size of 350 bp by shearing, then DNA

fragments were end-polished, A-tailed, and ligated with the NEBNext adapter for Illumina sequencing, and further PCR enriched by P5 and indexed P7 oligos. The PCR products were purified (AMPure XP system) and the resulting libraries were analyzed for size distribution by Agilent 2100 BioAnalyzer and quantified using real-time PCR. Since the genome sizes for the Troidini species was estimated to be about 320 Mb (*Ornithoptera*) and 340 Mb (*Troides*), Illumina 150 bp paired-end sequencing was run on a NovaSeq 6000 instrument to obtain about 32 and 34 Gb per sample corresponding to a genome depth-of-coverage of about 100x.

The quality and quantity of all RNAs were checked using Nanodrop, Qubit, and 1.0% agarose gel electrophoresis and sent to Novogene for library preparations. Messenger RNA was purified from total RNA using poly-T oligo-attached magnetic beads. After fragmentation, the first strand cDNA was synthesized using random hexamer primers followed by the second strand cDNA synthesis. The library was ready after end repair, A-tailing, adapter ligation, size selection, amplification, and purification. The library was checked with Qubit and real-time PCR for quantification and BioAnalyzer for size distribution detection. Quantified libraries have been pooled and sequenced on Illumina platforms, according to effective library concentration to a data amount of about 8 Gb per sample.

Assembly of reference genomes

For GridION sequencing, all fast5 files were basecalled using Guppy 5.0.15 (developed by ONT) using super-high accuracy mode and a quality control of 10 (min_score 10). Sequencing adapters were trimmed using Porechop 0.2.3 (https://github.com/rrwick/Porechop). Draft genome assemblies were performed with Flye 2.8.3 (Kolmogorov et al. 2019) with default options. Illumina reads were cleaned, filtered and paired with fastp 20.0 (Chen et al. 2018) using default options. Paired-end sequences were mapped on the Flye assembly using BWA 0.7.17 (Li 2013). Resulting SAM files were converted to sorted indexed BAM files with

SAMtools (Li et al. 2009). Flye draft assemblies were polished with two rounds of Pilon 1.24 (Walker et al. 2014) using this mapping information. Assembly statistics were then assessed using the gVolante2 platform (Nishimura et al. 2017) to retrieve the number and size of contigs, the presence, completeness and duplication of BUSCO genes of the Lepidoptera odb10 database (Manni et al. 2021). More information and statistics about quality of sequencing, assembly and polishing are displayed in Supplementary table S6, Supplementary Material online. Before submitting genomes assemblies to GenBank, we used BlobTools 1.1.1 (Laetsch and Blaxter 2017) set to the ncbi and diamond databases to check for possible contaminations. We found no evidence of artificial contamination coming from laboratory manipulation, but some contigs were clearly identified as belonging to exogenous organisms such as host plants and symbionts. We removed all contigs that were belonging to the Plant (subsequently identified as *Aristolochia* by BLAST), Microsporidia (unicellular fungal insect parasites) or Pseudomonadota (*Wolbachia*, *Enterobacter*) phylum. The FC563 assembly showed no evidence of contamination, while 83% of the total contigs removed were from FC561 (supplementary table S7, Supplementary Material online). We used the scaffold function of RagTag 2.1.0 (https://github.com/malonge/RagTag; Alonge et al. 2019; Alonge et al. 2021) to find the correspondence between *Papilio bianor* chromosomes and *O. alexandrae* FC563 assembly.

Genome annotation

We performed a full pipeline of annotations for the individuals FC560 and FC563. The pipeline was composed of the six following steps. First, we reconstructed the repeat sequences using RepeatModeler 2.0.1 (Flynn et al. 2020). The consensus sequences generated by RepeatModeler were blasted against the "reference transcriptome" database of UniProt (download in October 2021, https://www.uniprot.org/) using diamond blastx (Buchfink et al.

2015), and we excluded all the proteins that were not associated to repeat sequences from the consensus sequences. We then annotated the repeat sequences in the respective assemblies using RepeatMasker (Smit et al. 2013-2015) using both the Dfam libraries (Storer et al. 2021) setting the parameter "-species" on "Arthropoda" and the newly identified repeat sequences reconstructed using RepeatModeler. Second, we assembled the RNA-seq data by cleaning the reads with fastp 20.0 (Chen et al. 2018), mapped the read using HISAT2 (Kim et al. 2019) onto the reference genome, and we annotated the cDNA using StringTie (Petrea et al. 2015) producing a GTF file. The cDNA sequences were converted in fasta using the "gtf_genome_to_cdna_fasta.pl" script of TransDecoder (https://github.com/TransDecoder/TransDecoder). The RNA-seq data of all the individuals were mapped against the reference genome of FC563. Third, we ran MAKER 2.31.11 (Holt and Yandell 2011) using the information of the annotated repeat sequences and the cDNA sequences provided as "rm_gff" and "est" option in the control file of MAKER, respectively. We also used the proteins sequences of *Heliconius melpomene, Melitaea cinxia, Papilio machaon, Papilio xuthus, Papilio glaucus* provided as "protein" option in the control file of MAKER to help identify genes using homology information. Fourth, SNAP (Korf et al. 2004) and AUGUSTUS (Stanke et al. 2006) were used to produce gene prediction models from the first round of MAKER. BUSCO 5 (Simão et al. 2015) with options "--long" and "--augustus" and the Endopterygota database was used to produce the gene prediction model of AUGUSTUS. Fifth, we ran again MAKER using the annotation from the first round and the gene models of SNAP and AUGUSTUS. Sixth, the steps 3 and 4 were repeated using the second round of MAKER annotation to produce a third and final round of annotations. Finally, for individuals FC561 and FC562, we used Liftoff 1.6.3 (Shumate and Salzberg 2021) to map the annotation of FC560 on the two other assemblies.

578 ***Mitogenomic diversity***

579 Long reads of every individual were corrected with short reads using LoRDEC 0.9 (Salmela
580 and Rivals 2014). For each individual of *O. alexandrae*, the corrected long reads were mapped
581 with Minimap2 2.17 (Li 2018) on the reference mitogenome of *O. richmondia* from a previous
582 study (NC_037869.1, Condamine et al. 2018). The reads that mapped with the references were
583 filtered by quality via SAMtools (Li et al. 2009) ("view -q 30"). For each individual, a subset
584 of reads from 3.6 Mb to 6.3 Mb were created so that mitogenomes would have an expected
585 depth of coverage between 200x and 400x. We used Flye 2.8.3 to assemble the mitogenomes
586 and the resulting assemblies were given to MitoFinder 1.4 (Allio et al. 2020) to annotate (gene,
587 tRNA, rRNA) and extract genes (gene, rRNA). The cleaned short-read data were also directly
588 given to MitoFinder to produce annotated mitogenomes based on the short-read data only. As
589 these mitogenomes were of slightly better quality (better annotation, presence of the complete
590 or nearly complete D-loop), which were those submitted to GenBank (Accession numbers
591 OQ59006-OQ59009). A nucleotide alignment was produced with MAFFT 7.310 (Katoh and
592 Standley 2013) after having manually adjusted the sequences due to circularization. We used
593 Seaview 4.7 (Gouy et al. 2010) to visualize the four whole mitochondrial genome alignment
594 and to count pairwise differences using "Statistics" of Seaview, and the mean of this pairwise
595 distance was calculated, and ultimately divided by the alignment length (supplementary table
596 S3, Supplementary Material online). The same steps were carried out with the COI
597 mitochondrial gene only.
598
599 ***Nuclear heterozygosity of Troidini***

600 For *O. alexandrae*, we selected FC563 as the reference assembly, as it has the highest N50,
601 mean coverage and BUSCO score, the lowest number of contigs, and has no contamination
602 (Table 1). Genomes of *O. priamus* and *T. oblongomaculatus* were their own reference. The

603 corrected reads (LoRDEC, see *Mitogenomic diversity* section) of every individual were
604 mapped on their reference genome using "-a" option of Minimap2 2.17 (Li 2018). We used
605 SAMtools to compress, sort and index these mappings. SNP calling was performed with
606 Longshot 0.4.1 (Edge and Bansal 2019), using a threshold of 15x minimum and 150x maximum
607 for the depth of coverage (minimal quality of 20; default quality) and applying a
608 transition/transversion rate for genotype prior estimation (ts_tv_ratio) of 2.0 (Edge and Bansal
609 2019). SNPs with a quality below 200 were excluded. All positions, SNPs and homozygous,
610 must be contained within the coverage thresholds, otherwise they were considered as
611 ambiguous. As the quality of phasing may be important for population genomics and
612 demographic analyses, we checked the average size of phasing from Longshot. The average
613 length of the phased blocks is 435 kb, and a haplotype N50 of 1.7 Mb (Supplementary table
614 S8, Supplementary Material online for phasing statistics per individual). In addition, to ensure
615 that our heterozygosity estimates did not depend on the data and method (Bentley and
616 Armstrong 2022), we also calculated the heterozygosity rate based on the short reads data to
617 evaluate the robustness of our results. Illumina cleaned reads were mapped on references using
618 the speedseq pipeline (Chiang et al. 2015) that relies on BWA 0.7.17 (Li 2013). We excluded
619 the so-called discordant and splitter reads and the reads with mapping quality below 30.
620 Genotype calling was performed using FreeBayes 1.3.1 (Garrison and Marth 2012) set with the
621 "--use-best-n-alleles 4" option, and the same coverage threshold as for long reads data. SNPs
622 with quality scores below 50 and out of these coverage thresholds were excluded. Homozygous
623 positions were also selected based on the same coverage threshold and considered ambiguous
624 otherwise. For ONT and Illumina data, heterozygosity was computed as the number of SNPs
625 divided by the total number of sites excluding ambiguous positions. Using the above criteria,
626 heterozygosity levels were similar between ONT and Illumina data (supplementary table S4,
627 Supplementary Material online). To compare our results with the values of Mackintosh et al.

(2019), heterozygosity was also computed using only four-fold degenerate sites, as identified by a custom script using BIO++ library (Guéguen et al. 2013).

Estimation of the demographic history and effective population size

We used a Sequential Markovian Coalescent (SMC) model (McVean and Cardin 2005, e.g. PSMC, Li and Durbin 2011; MSMC2, Schiffels and Wang 2020) to estimate the ancestral effective population size (*Ne*) trajectory of the studied Troidini species.

The SMC model requires calibrations, in particular a value of mutation rate. We estimated this rate based on synonymous mutations by selecting the four-fold degenerate sites of the third codon positions of BUSCO genes from the odb10 lepidopteran database. We retrieved the set of fasta nucleotide sequences using BUSCOMP 0.13.0 (Edwards 2019) on local runs of BUSCO 5 (odb10_Lepidoptera) of the six studied individuals and we considered only the genes that contained all individuals, which corresponded to 5,127 genes (~97% of the lepidopteran gene database). Assuming that these mutations are neutral, we applied the formula $D = 2 \times T \times \mu$ where D is the genetic divergence between two species, T is the divergence in millions of years and μ is the mutation rate per million years (Kimura 1983; Birky and Walsh 1988). Here, we chose *O. priamus* as the divergent species of *O. alexandrae*, and set T=12.03166 Ma (median value of the divergence time between these two species, with 95% credibility interval = 7.9662–16.7068; *sensu* Allio et al. 2021). To estimate D, we split the 5,127 genes into six bins based on the GC-content and we estimated the branch lengths from the six corresponding trees inferred by IQ-TREE 1.6.12 (Nguyen et al. 2015) with a GTR+Γ6 substitution model. The divergence D ranged from 0.043 to 0.052 for the lowest GC-content bins to the highest, respectively. We then took an average D (= 0.0475) between the two *Ornithoptera* species to obtain a mean value of μ equal to 1.9740e-09 mutations per site per year. As *O. alexandrae* highland population produces one generation per year while the

lowland population produces two (Mitchell et al. 2016), we set the generation time to one and a half generations per year, therefore the mutation rate μ was estimated at 1.3160e-09 mutations per site per generation, which is at the lowest end of the range estimated for *Heliconius* (μ = 1.3 - 5.5e-09; Keightley et al. 2015).

It has been shown that SMC models do not perform well when the ratio of recombination rate r over mutation rate μ becomes greater than one (Sellinger et al. 2021). Assuming *O. alexandrae* genome is 325 Mb long, distributed in 30 chromosomes and that there is a single crossover per tetrad per male meiosis, the recombination rate would be r = 2.7e-8. An analysis of the nymphalid *Vanessa cardui* (Shipilina et al. 2022) estimated an average r between 3.81e-8 and 4.05e-8 with substantial inter-chromosomal variation, meaning that the average recombination rate of *O. alexandrae* is an order of magnitude greater than its mutation rate at least.

To investigate if this parameter range was a problem in SMC analyses, we used msprime 1.2.0 (Kelleher et al. 2016) to produce ten simulated datasets of a stable demographic history scenario with *Ne* = 100,000 under a r = 1e-8 and a μ = 1.316e-9 on 30 chromosomes of 10 Mb each. We ran the Multiple Sequentially Markovian Coalescent (MSMC) model implemented in the MSMC2 software (https://github.com/stschiff/msmc2; Schiffels and Wang 2020) with default options (i.e. *-rhoOverMu = 0.25*) to test whether the inferred demography was recovered stable. We ran the same data by setting an initial value *-rhoOverMu = 10*.

We applied the MSMC2 model on real data, first for each Troidini species (FC563 was selected for *O. alexandrae*). We used the VCF files generated using Longshot (as described in the *Nuclear heterozygosity of Troidini* section) and created the so-called "mask file" for each individual based on the depth of coverage thresholds of >20x and <150x using a custom python script. These files were then combined using the "generate_multihetsep.py" of MSMC2 to generate "multihetsep.txt" input files (https://github.com/stschiff/msmc-

tools/blob/master/msmc-tutorial/guide.md). MSMC2 was run using default parameters (especially the initial value of the ratio [r / μ] *-rhoOverMu = 0.25)*, and with *-rhoOverMu = 10* (See Supplementary figure S2, Supplementary Material online for a comparison of the results with both options).

By applying the same methods, we ran MSMC2 on each individual of *O. alexandrae* and then applied the model on the two populations (both composed by two genomes). MSMC2 was run using default parameters, or with *-rhoOverMu = 10*. The "-I" option was used to consider relevant haplotypes depending on the analyses (i.e., single individuals, two populations and cross coalescence rate between populations). For each analysis on real data, we generated 10 bootstraps per individual using the multihetsep_bootstrap.py script available at: https://github.com/stschiff/msmc-tools. We generated all graphs with the R package ggplot2 (Wickham 2016) by considering a generation time of 0.5 for *O. priamus* and *T. oblongomaculatus* (two generations per year), 0.75 for *O. alexandrae* (one and a half generations per year) and μ = 1.316e-9 for every individual.Finally, we relied on the ABC framework implemented in DILS (Fraïsse et al. 2021) to test several scenarios of divergence between populations. Alternative methods like ∂a∂i (Gutenkunst et al. 2009) or FastSimCoal (Excoffier and Foll 2011) could not be implemented because of no modeling of *Ne* through time and sample size limitations of our dataset to compute site-frequency spectrum, respectively. DILS takes into account linkage information that is informative about past demography (Fraïsse et al. 2021). To fit DILS on our data, we randomly selected 5,000 windows of 4,000 bp (2 Mb in total) because the coalescence program is time-consuming to simulate large chunks of chromosomes with recombination. The analysis was replicated four times to evaluate variability and reproducibility of the ABC inferences. DILS implements a pipeline that selects the best-fitting demographic model by comparing models with variations in *Ne* and migration among loci allowing to consider linked selection and alleles that could be

selected against during hybridization (Fraïsse et al. 2021). The four demographic scenarios tested include: strict isolation (SI), isolation with migration (IM), ancient migration (AM), and secondary contact (SC). DILS used a random forest method (Pudlo et al. 2016) to select the best model and estimate posterior parameters' distributions using rejection and neural network methods implemented in the R package abc (Csilléry et al. 2012).

Population structure

The population differentiation due to genetic structure was estimated with the nucleotide diversity and fixation index (F_{ST}) that were computed using seq_stat_2pop (https://github.com/benoitnabholz/seq_stat_2pop) using Bio++ library (Guéguen et al. 2013). The seq_stat_2pop program uses fasta sequences as input such that VCF files were converted into fasta sequences using a custom python program using coverage information for the homozygous sites as explained above (see *Nuclear heterozygosity of Troidini* section). Nucleotide diversity (π-diversity) was computed as the mean pairwise divergence between pairs of chromosomes. F_{ST} was computed using the nucleotide sequence as: $F_{ST} = 1 - \pi_{intra} / \pi_{total}$; where π_{intra} is the mean nucleotide diversity of the two populations ($\pi_{intra} = (\pi_{highland}+\pi_{lowland}) / 2$) and π_{total} is the nucleotide diversity computed using all individuals (Nei 1982). F_{ST} and nucleotide diversity were computed on non-sex-related contigs >500 kb in length, over sliding windows of 100 kb with 50 kb overlapping regions (windows of 100 kb and overlapping regions of 100 kb were tested and led to similar results). It is considered that a F_{ST} value greater than 0.15 is significant in differentiating populations (Frankham et al. 2010).

Acknowledgments

We thank two anonymous reviewers who provided very relevant and constructive comments to improve the study. This project has received funding from the European Research Council

(ERC) under the European Union's Horizon 2020 research and innovation programme (project GAIA, agreement no. 851188). This project has been generously supported by the Swallowtail & Birdwing Butterfly Trust, and we thank the Trust's Chair, N. Mark Collins, for his continuous advice and encouragement. We thank Henry S. Barlow and Ian Orrell who helped with the organization of the field work as well as with the export and CITES permits. We thank Conwell Nukara for field work assistance in the Managalas Plateau. This study also benefited from lab facilities near Popondetta funded by the Sime Darby Foundation. CITES permits were obtained for *Ornithoptera alexandrae* (FR1903400125-I, 019418, delivered October 23, 2019), *Ornithoptera priamus* (FR1903400124-I, 019429, delivered November 1, 2019), *Troides oblongomaculatus* (FR1903400124-I, 019429, delivered November 1, 2019).

Author Contributions

F.L.C. and D.B. conceived and supervised the project. F.L.C. and D.B. collected the samples. E.C. and M.-K.T. carried out the molecular experiments. E.L.R. and B.N. performed the bioinformatic analyses. E.L.R., B.N., and F.L.C. discussed the results. E.L.R. and F.L.C. wrote the draft manuscript and B.N., M.-K.T and D.B. made comments. All authors have read and approved the final manuscript.

Data Availability

The birdwing genomes, mitogenomes and sequencing data in the present study, including Nanopore, Illumina, and RNA data are available from the Genome database and Sequence Read Archive (SRA) under the Bioproject accession number PRJNA938052, with the corresponding BioSamples accession numbers FC560: SAMN33424250; FC561: SAMN33424251; FC562: SAMN33424252; FC563: SAMN33424253; FC565: SAMN33424254; FC569: SAMN33424255.

753

754 **References**

755 Ackery PR. 1997. The Natural History Museum collection of *Ornithoptera* (birdwing)
756 butterflies (Lepidoptera: Papilionidae). *Biol Curat.* **8**:11-17.

757 Allio R, Donega S, Galtier N, Nabholz B. 2017. Large variation in the ratio of mitochondrial
758 to nuclear mutation rate across animals: implications for genetic diversity and the use of
759 mitochondrial DNA as a molecular marker. *Mol Biol Evol.* **34**:2762-2772.

760 Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz B, Delsuc F. 2020.
761 MitoFinder: efficient automated large-scale extraction of mitogenomic data in target
762 enrichment phylogenomics. *Mol Ecol Res.* **20**:892-905.

763 Allio R, Nabholz B, Wanke S, Chomicki G, Pérez-Escobar OA, Cotton AM, Clamens AL,
764 Kergoat GJ, Sperling FAH, Condamine FL. 2021. Genome-wide macroevolutionary
765 signatures of key innovations in butterflies colonizing new host plants. *Nat Commun.*
766 **12**:354.

767 Alonge M, Lebeigle L, Kirsche M, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S.
768 unpublished data. Automated assembly scaffolding elevates a new tomato system for
769 high-throughput genome editing. *bioRxiv.*
770 https://www.biorxiv.org/content/10.1101/2021.11.18.469135v1, last accessed
771 November 9, 2022.

772 Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz
773 MC. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes.
774 *Genome Biol.* **20**:224.

775 Barrows TT, Hope GS, Prentice ML, Fifield LK, Tims SG. 2011. Late Pleistocene glaciation
776 of the Mt Giluwe volcano, Papua New Guinea. *Quatern. Sci Rev.* **30**: 2676-2689.

777 Bentley BP, Armstrong EE. 2022. Good from far, but far from good: The impact of a reference
778 genome on evolutionary inference. *Mol Ecol Res.* **22**:12–14.
779 Birky CW Jr, Walsh JB. 1988. Effects of linkage on rates of molecular evolution. *Proc Natl*
780 *Acad Sci.* **85**:6414–6418.
781 Böhm M. 2018. *Ornithoptera alexandrae*. The IUCN Red List of Threatened Species 2018:
782 https://dx.doi.org/10.2305/IUCN.UK.2018-1.RLTS.T15513A88565197.en.
783 Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND.
784 *Nat Methods* **12**:59–60.
785 Buchwalder K, Samain MS, Sankowsky G, Neinhuis C, Wanke S. 2014. Nomenclatural
786 updates of *Aristolochia* subgenus *Pararistolochia* (Aristolochiaceae). *Austral Syst Bot.*
787 **27**:48-55.
788 Buffalo V. 2021. Quantifying the relationship between genetic diversity and population size
789 suggests natural selection cannot explain Lewontin's Paradox. *eLife* **10**:e67509.
790 Burns JM, Janzen DH, Hajibabaei M, Hallwachs, W, Hebert PDN. 2008. DNA barcodes and
791 cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservacion
792 Guanacaste, Costa Rica. *Proc Natl Acad Sci.* **105**:6350-6355.
793 Burri R. 2017. Interpreting differentiation landscapes in the light of long-term linked selection.
794 *Evol Lett.* **1**:118–131.
795 Chappell J, Polach H. 1991. Post-glacial sea-level rise from a coral record at Huon Peninsula,
796 Papua New Guinea. *Nature* **349**:147-149.
797 Chattopadhyay B, Garg KM, Soo YJ, Low GW, Frechette JL, Rheindt FE. 2019. Conservation
798 genomics in the fight to help the recovery of the critically endangered Siamese crocodile
799 *Crocodylus siamensis*. *Mol Ecol.* **28**:936-950.
800 Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor.
801 *Bioinformatics* **34**:i884-i890.

802 Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, Marth GT, Quinlan
803 AR, Hall IM. 2015. SpeedSeq: ultra-fast personal genome analysis and interpretation.
804 *Nat Methods* **12**:966–968.
805 Collins NM, Morris MG. 1985. Threatened swallowtail butterflies of the world. The IUCN Red
806 Data Book, Cambridge.
807 Condamine FL, Nabholz B, Clamens AL, Dupuis JR, Sperling FAH. 2018. Mitochondrial
808 phylogenomics, the origin of swallowtail butterflies, and the impact of the number of
809 clocks in Bayesian molecular dating. *Syst Entomol.* **43**:460-480.
810 Cong Q, Borek D, Otwinowski Z, Grishin NV. 2015. Tiger swallowtail genome reveals
811 mechanisms for speciation and caterpillar chemical defense. *Cell Rep*. **10**:910-919.
812 Csilléry K, François O, Blum MG. 2012. abc: an R package for approximate Bayesian
813 computation (ABC). *Meth Ecol Evol.* **3**:475–479.
814 Delmore KE, Lugo Ramos JS, Van Doren BM, Lundberg M, Bensch S, Irwin DE, Liedvogel
815 M. 2018. Comparative analysis examining patterns of genomic differentiation across
816 multiple episodes of population divergence in birds. *Evol Lett.* **2**:76–87.
817 Díez-del-Molino D, Sánchez-Barreiro F, Barnes I, Gilbert MTP, Dalén L. 2018. Quantifying
818 Temporal Genomic Erosion in Endangered Species. *Trends Ecol Evol.* **33**:176–185.
819 Dincă V, Dapporto L, Somervuo P, Vodă R, Cuvelier S, Gascoigne-Pees M, Huemer P,
820 Mutanen M, Hebert PDN, Vila, R. 2021. High resolution DNA barcode library for
821 European butterflies reveals continental patterns of mitochondrial genetic diversity.
822 *Commun Biol.* **4**:315.
823 Dussex N, Van Der Valk T, Morales, HE, Wheat CW, Díez-del-Molino D, Von Seth J, Foster
824 Y, Kutschera VE, Guschanski K, Rhie K, *et al.* 2021. Population genomics of the
825 critically endangered kākāpō. *Cell Genomics* **1**:100002.

826 Edge P, Bansal V. 2019. Longshot enables accurate variant calling in diploid genomes from
827 single-molecule long read sequencing. *Nat Commun.* **10**:4660.
828 Edwards RJ. 2019. BUSCOMP: BUSCO compilation and comparison–Assessing
829 completeness in multiple genome assemblies. *F1000Research* **8**:995.
830 Ellegren H, Galtier N. 2016. Determinants of genetic diversity. *Nat Rev Genet.* **17**:422-433.
831 Excoffier L, Foll M. 2011. Fastsimcoal: a continuous-time coalescent simulator of genomic
832 diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* **27**:1332-
833 1334.
834 Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020.
835 RepeatModeler2 for automated genomic discovery of transposable element families.
836 *Proc Natl Acad Sci.* **117**:9451–9457.
837 Formenti G, Theissinger K, Fernandes C, Bista I, Bombarely A, Bleidorn C, Ciofi C, Crottini
838 A, Godoy JA, Höglund J, *et al*. 2022. The era of reference genomes in conservation
839 genomics. *Trends Ecol Evol.* **37**:197–202.
840 Fraïsse C, Popovic I, Mazoyer C, Spataro B, Delmotte S, Romiguier J, Loire E, Simon A,
841 Galtier N, Duret L, *et al.* 2021. DILS: Demographic inferences with linked selection by
842 using ABC. *Mol Ecol Res.* **21**:2629-2644.
843 Frankham R, Ballou JD, Briscoe DA. 2010. Introduction to conservation genetics. Cambridge
844 University Press, UK.
845 Fuller ZL, Mocellin VJ, Morris LA, Cantin N, Shepherd J, Sarre L, Peng J, Liao Y, Pickrell
846 PA, Matz M, et al. 2020. Population genetics of the coral *Acropora millepora*: Toward
847 genomic prediction of bleaching. *Science.* **369**:eaba4674.
848 García-Berro A, Talla V, Vila R, Wai HK, Shipilina D, Chan KG, Pierce NE, Backström N,
849 Talavera G. 2023. Migratory behavior is positively associated with genetic diversity in
850 butterflies. *Mol Ecol.* **32**:560-574.

Garner BA, Hand BK, Amish SJ, Bernatchez L, Foster JT, Miller KM, Morin PA, Narum SR, O'Brien SJ, Roffler G, *et al*. 2016. Genomics in conservation: case studies and bridging the gap between data and application. *Trends Ecol Evol.* **31**:81-83.

Garrison E, Marth G. unpublished data. Haplotype-based variant detection from short-read sequencing. *arXiv*. https://doi.org/10.48550/arXiv.1207.3907 last accessed November 9, 2022.

Global Volcanism Program, 2022. Lamington (253010) in Volcanoes of the World (v. 5.0.0; 1 Nov 2022). Distributed by Smithsonian Institution, compiled by Venzke, E. https://doi.org/10.5479/si.GVP.VOTW5-2022.5.0

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol.* **27**:221-224.

Guéguen L, Gaillard S, Boussau B, Gouy M, Groussin M, Rochette NC, Bigot T, Fournier D, Pouyet F, Cahais V, *et al.* 2013. Bio++: Efficient extensible libraries and tools for computational molecular evolution. *Mol Biol Evol.* **30**:1745–1750.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**:e1000695.

Haugum J, Low AM. 1979. A monograph of the birdwing butterflies. The systematics of *Ornithoptera*, *Troides* and related genera. Vol 1, *Ornithoptera*. Scandinavian Science Press, Klampenborg, 308 pp.

Hayward MW, Castley JG. 2018. Editorial: triage in conservation. *Front Ecol Evol*. **5**:168.

Hazzouri KM, Sudalaimuthuasari N, Kundu B, Nelson D, Al-Deeb MA, Le Mansour A, Spencer JJ, Desplan C, Amiri KMA. 2020. The genome of pest *Rhynchophorus*

ferrugineus reveals gene families important at the plant-beetle interface. *Commun Biol.* **3**:323.

He JW, Zhang R, Yang J, Chang Z, Zhu LX, Lu SH, Xie FA, Mao JL, Dong ZW, Liu GC, *et al*. 2022. High-quality reference genomes of swallowtail butterflies provide insights into their coloration evolution. *Zool Res.* **43**:367-379.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc Natl Acad Sci.* **101**:14812-14817.

Heckenhauer J, Frandsen PB, Sproul JS, Li Z, Paule J, Larracuente AM, Maughan PJ, Barker MS, Schneider JV, Stewart RJ, *et al*. 2022. Genome size evolution in the diverse insect order Trichoptera. *GigaScience.* **11**:giac011.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**:491.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* **30**:772-780.

Kebaïli C, Sherpa S, Rioux D, Després L. 2022. Demographic inferences and climatic niche modelling shed light on the evolutionary history of the emblematic cold-adapted Apollo butterfly at regional scale. *Mol Ecol.* **31**:448-466.

Keightley PD, Pinharanda A, Ness RW, Simpson F, Dasmahapatra KK, Mallet J, Davey JW, Jiggins CD. 2015. Estimation of the spontaneous mutation rate in *Heliconius melpomene*. *Mol Biol Evol.* **32**:239-243.

Kelleher J, Etheridge AM, McVean G. 2016. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput Biol*. **12**:e1004842.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* **37**:907–915.

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge University Press, UK.

Koh LP, Sodhi NS, Brook BW. 2004. Ecological correlates of extinction proneness in tropical butterflies. *Cons Biol.* **18**:1571-1578.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotech.* **37**:540-546.

Korf I. 2004 Gene finding in novel genomes. *BMC Bioinformatics* **5**:59.

Laetsch DR, Blaxter ML. 2017. BlobTools: Interrogation of genome assemblies. *F1000Research* **6**:1287.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* **475**:493–496.

Li H. 2013. Unpublished data. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv.* https://doi.org/10.48550/arXiv.1303.3997, last accessed November 9, 2022.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094-3100.

Li H. Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**:2078-2079.

Liu G, Chang Z, Chen L, He J, Dong Z, Yang J, Lu S, Zhao R, Wan W, Ma G. 2020. Genome size variation in butterflies (Insecta, Lepidoptera, Papilionoidea): a thorough phylogenetic comparison. *Syst Entomol.* **45**:571–582.

Lu S, Yang J, Dai X, Xie F, He J, Dong Z, Mao J, Liu G, Chang Z, Zhao R, *et al.* 2019. Chromosomal-level reference genome of Chinese peacock butterfly (*Papilio bianor*) based on third-generation DNA sequencing and Hi-C analysis. *GigaScience* **8**:giz128.

Lynch M, Ackerman MS, Gout JF, Long H, Sung W, Thomas WK, Foster PL. 2016. Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet.* **17**:704-714.

925 Mackintosh A, Laetsch DR, Hayward A, Charlesworth B, Waterfall M, Vila R, Lohse K. 2019.
926 The determinants of genetic diversity in butterflies. *Nat Commun.* **10**:3466.
927 Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel
928 and streamlined workflows along with broader and deeper phylogenetic coverage for
929 scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol.* **38**:4647-4654.
930 Manthey JD, Girón JC, Hruska JP. 2022. Impact of host demography and evolutionary history
931 on endosymbiont molecular evolution: A test in carpenter ants (genus *Camponotus*) and
932 their *Blochmannia* endosymbionts. *Ecol Evol.* **12**:e9026.
933 Mattila N, Kaitala V, Komonen A, Kotiaho JS, Päivinen J. 2006. Ecological determinants of
934 distribution decline and risk of extinction in moths. *Cons Biol.* **20**:1161-1168.
935 McVean GA, Cardin NJ. 2005. Approximating the coalescent with recombination. *Phil Trans*
936 *R Soc B.* **360**:1387-1393.
937 Meek AS. 1906. Letter to Karl Jordan dated 3rd Feb 1906 from "Above "Biagi" (5000 ft) Head
938 of Mambare River", Papua New Guinea. Meek letter 155, BMNH Archives.
939 Mikheyev AS, Zwick A, Magrath MJ, Grau ML, Qiu L, Su YN, Yeates D. 2017. Museum
940 genomics confirms that the Lord Howe Island stick insect survived extinction. *Curr Biol.*
941 **27**:3157-3161.
942 Mitchell DK, Dewhurst CF, Tennent WJ, Page WW. 2016. Queen Alexandra's birdwing
943 butterfly *Ornithoptera alexandrae* (Rothschild, 1907): A review and conservation
944 proposals. Kuala Lumpur: Southdene Sdn Bhd.
945 Morin PA, Foote AD, Hill CM, Simon-Bouhet B, Lang AR, Louis M. 2018. SNP discovery
946 from single and multiplex genome assemblies of non-model organisms. In: Head SR,
947 Ordoukhanian P, Salomon DR, editors. Next Generation Sequencing. New York (NY):
948 Springer New York. p. 113-144

949 Morin PA, Archer FI, Avila CD, Balacco JR, Bukhman YV, Chow W, Fedrigo O, Formenti G,
950 Fronczek JA, Fungtammasan A, *et al.* 2021. Reference genome and demographic history
951 of the most endangered marine mammal, the vaquita. *Mol Ecol Res.* **21**:1008-1020.
952 Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, Ellegren H. 2015. Temporal dynamics of
953 avian populations during Pleistocene revealed by whole-genome sequences. *Curr Biol.*
954 **25**:1375-1380.
955 Nakae M. 2021. *Papilionidae of the World*. Tokyo: Roppon-Ashi Entomological Books.
956 Nei M. 1982. Evolution of human races at the gene level. In: Bonné-Tamir B, editor. Human
957 Genetics, Part A: The unfolding Genome. New York (NY): Alan R. Liss. p. 167–181.
958 Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective
959 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.*
960 **32**:268-274.
961 Nishimura O, Hara Y, Kuraku S. 2017. gVolante for standardizing completeness assessment
962 of genome and transcriptome assemblies. *Bioinformatics* **33**:3635-3637.
963 Palash A, Paul S, Resha SK, Khan MK. 2022. Body size and diet breadth drive local extinction
964 risk in butterflies. *Heliyon* **8**:e10290.
965 Parsons MJ. 1992. The world's largest butterfly endangered: the ecology, status and
966 conservation of *Ornithoptera alexandrae* (Lepidoptera: Papilionidae). *Trop Lepido Res.*
967 **3**:33-60.
968 Parsons MJ. 1996. New species of *Aristolochia* and *Pararistolochia* (Aristolochiaceae) from
969 Australia and New Guinea. *Botanical Journal of the Linnean Society,* **120**:199-238.
970 Parsons MJ. 1999. The butterflies of Papua New Guinea: their systematics and biology.
971 London: Academic Press.

Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* **33**:290–295.

Petersen M, Armisén D, Gibbs RA, Hering L, Khila A, Mayer G, Richards S, Niehuis O, Misof B. 2019. Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. *BMC Ecol Evol.* **19**:11.

Podsiadlowski L, Tunström K, Espeland M, Wheat CW. 2021. The genome assembly and annotation of the Apollo butterfly *Parnassius apollo*, a flagship species for conservation biology. *Genome Biol Evol.* **13**:evab122.

Pudlo P, Marin J-M, Estoup A, Cornuet J-M, Gautier M, Robert CP. 2016. Reliable ABC model choice via random forests. *Bioinformatics* **32**:859–866.

Quinn RM, Gaston KJ, Roy DB. 1997. Coincidence between consumer and host occurrence: macrolepidoptera in Britain. *Ecol Entomol*. **22**:197–208.

Robinson JA, Brown C, Kim BY, Lohmueller KE, Wayne R.K. 2018. Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. *Curr Biol.* **28**:3487-3494.

Robinson JA, Bowie RC, Dudchenko O, Aiden EL, Hendrickson SL, Steiner CC, Ryder OA, Mindell DP, Wall JD. 2021. Genome-wide diversity in the California condor tracks its prehistoric abundance and decline. *Curr Biol*. **31**:2939-2946.

Robinson JA, Kyriazis CC, Nigenda-Morales SF, Beichman AC, Rojas-Bracho L, Robertson KM, Fontaine MC, Wayne RK, Lohmueller KE, Taylor BL, *et al*. 2022. The critically endangered vaquita is not doomed to extinction by inbreeding depression. *Science* **376**:635-639.

995 Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dernat R,
996 Duret L, Faivre N, *et al.* 2014. Comparative population genomics in animals uncovers
997 the determinants of genetic diversity. *Nature* **515**:261-263.
998 Salmela L, Rivals E. 2014. LoRDEC: accurate and efficient long read error correction.
999 *Bioinformatics* **30**:3506-3514.
1000 Sánchez-Bayo F, Wyckhuys KA. 2019. Worldwide decline of the entomofauna: A review of
1001 its drivers. *Biol Cons.* **232**:8-27.
1002 Schiffels S, Wang K. 2020. MSMC and MSMC2: The Multiple Sequentially Markovian
1003 Coalescent. In: Dutheil JY, editor. Statistical Population Genomics. New York (NY):
1004 Springer New York. p. 147–166.
1005 Sellinger TPP, Abu Awad D, Moest M, Tellier A. 2020. Inference of past demography,
1006 dormancy and self-fertilization rates from whole genome sequence data. *PLoS Genet.*
1007 **16**:e1008698
1008 Sellinger TPP, Abu Awad D, Tellier A. 2021. Limits and convergence properties of the
1009 sequentially Markovian coalescent. *Mol Ecol Res.* **21**:2231-2248.
1010 Shipilina D, Näsvall K, Höök L, Vila R, Talavera G, Backström N. 2022. Linkage mapping
1011 and genome annotation give novel insights into gene family expansions and regional
1012 recombination rate variation in the painted lady (*Vanessa cardui*) butterfly. *Genomics*
1013 **114**:110481.
1014 Shumate A, Salzberg SL. 2021. Liftoff: accurate mapping of gene annotations. *Bioinformatics*
1015 **37**:1639–1643.
1016 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO:
1017 assessing genome assembly and annotation completeness with single-copy orthologs.
1018 *Bioinformatics* **31**:3210–3212.

1019 Smit AFA, Hubley R, Green P. 2013-2015. RepeatMasker Open-4.0.
1020 http://www.repeatmasker.org.
1021 Sproul JS, Hotaling S, Heckenhauer J, Powell A, Larracuente AM, Kelley JL, Pauls SU,
1022 Frandsen PB. unpublished data. Repetitive elements in the era of biodiversity genomics:
1023 insights from 600+ insect genomes. *bioRxiv*. https://doi.org/10.1101/2022.06.02.494618,
1024 last accessed November 9, 2022.
1025 Stapley J, Feulner PG, Johnston SE, Santure AW, Smadja CM. 2017. Variation in
1026 recombination frequency and distribution across eukaryotes: patterns and processes. *Phil*
1027 *Trans R Soc B*. **372**:20160455.
1028 Stanke M, Tzvetkova A, Morgenstern B. 2006. AUGUSTUS at EGASP: using EST, protein
1029 and genomic alignments for improved gene prediction in the human genome. *Genome*
1030 *Biol*. **7**:S11.
1031 Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. 2021. The Dfam community resource of
1032 transposable element families, sequence models, and genome annotations. *Mobile DNA*
1033 **12**:2.
1034 Straatman R. 1971. The life history of the *Ornithoptera alexandrae* Rothschild. *J Lepido Soc.*
1035 **25**:58-64.
1036 Teixeira H, Salmona J, Arredondo A, Mourato B, Manzi S, Rakotondravony R, Mazet O,
1037 Chikhi L, Metzger J, Radespiel, U. 2021. Impact of model assumptions on demographic
1038 inferences: the case study of two sympatric mouse lemurs in northwestern Madagascar.
1039 *BMC Ecol Evol*. **21**:1-18.
1040 Tennent WJ. 2021. The man who shot butterflies. Oxfordshire: Storm Entomological
1041 Publications.

1042 Tilak MK, Allio R, Delsuc F. 2020. An optimized protocol for sequencing mammalian roadkill
1043 tissues with Oxford Nanopore Technology (ONT) V1. *Protocols.io.*
1044 https://dx.doi.org/10.17504/protocols.io.6bthann
1045 Tine M, Kuhl H, Gagnaire P-A, Louro B, Desmarais E, Martins RST, Hecht J, Knaust F,
1046 Belkhir K, Klages S, *et al.* 2014. European sea bass genome and its variation provide
1047 insights into adaptation to euryhalinity and speciation. *Nat Commun.* **5**:5770.
1048 Tunstall T, Kock R, Vahala J, Diekhans M, Fiddes I, Armstrong J, Paten B, Ryder OA, Steiner
1049 CC. 2018. Evaluating recovery potential of the northern white rhinoceros from
1050 cryopreserved somatic cells. *Genome Res.* **28**:780-788.
1051 Van Der Valk T, Gonda CM, Silegowa H, Almanza S, Sifuentes-Romero I, Hart TB, Hart JA,
1052 Detwiler KM, Guschanski K. 2020. The genome of the endangered dryas monkey
1053 provides new insights into the evolutionary history of the vervets. *Mol Biol Evol.* **37**:183-
1054 194.
1055 Waldvogel AM, Wieser A, Schell T, Patel S, Schmidt H, Hankeln T, Feldmeyer B, Pfenninger
1056 M. 2018. The genomic footprint of climate adaptation in *Chironomus riparius*. *Mol Ecol.*
1057 **27**:1439-1456.
1058 Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q,
1059 Wortman J, Young SK, *et al*. 2014. Pilon: an integrated tool for comprehensive microbial
1060 variant detection and genome assembly improvement. *PLoS One* **9**:e112963.
1061 Walton W, Stone GN, Lohse K. 2021. Discordant Pleistocene population size histories in a
1062 guild of hymenopteran parasitoids. *Mol Ecol.* **30**:4538-4550.
1063 Wang K, Mathieson I, O'Connell J, Schiffels S. 2020. Tracking human population structure
1064 through time from whole genome sequences. *PLoS Genet.* **16**:e1008552.
1065 Wang S, Teng D, Li X, Yang P, Da W, Zhang Y, Zhang Y, Liu G, Zhang X, Wan W, *et al.*
1066 2022. The evolution and diversification of oakleaf butterflies. *Cell.* **185**:3138-3152.

1067 Wang Y, Obbard DJ. 2023. Experimental estimates of germline mutation rate in eukaryotes: a
1068 phylogenetic meta-analysis. *bioRxiv*. https://doi.org/10.1101/2023.01.24.525323, last
1069 accessed February 11, 2023.
1070 Warren WC, Boggs TE, Borowsky R, Carlson BM, Ferrufino E, Gross JB, Hillier L, Hu Z,
1071 Keene AC, Kenzior A, *et al.* 2021. A chromosome-level genome of *Astyanax mexicanus*
1072 surface fish for comparing population-specific genetic differences contributing to trait
1073 evolution. *Nat Commun.* **12**:1447.
1074 Westbury MV, Hartmann S, Barlow A, Wiesel I, Leo V, Welch R, Parker DM, Sicks F, Ludwig
1075 A, Dalén L, *et al.* 2018. Extended and continuous decline in effective population size
1076 results in low genomic diversity in the world's rarest hyena species, the brown hyena.
1077 *Mol Biol Evol.* **35**:1225-1237.
1078 Wickham H. 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York
1079 Available from: http://ggplot2.tidyverse.org
1080 Wilfert L, Gadau J, Schmid-Hempel P. 2007. Variation in genomic recombination rates among
1081 animal taxa and the case of social insects. *Heredity*. **98**:189-197.
1082 You M, Ke F, You S, Wu Z, Liu Q, He W, Baxter SW, Yuchi Z, Vasseur L, Gurr GM, *et al.*
1083 2020. Variation among 532 genomes unveils the origin and evolutionary history of a
1084 global insect herbivore. *Nat Commun.* **11**:2321.
1085

Figures

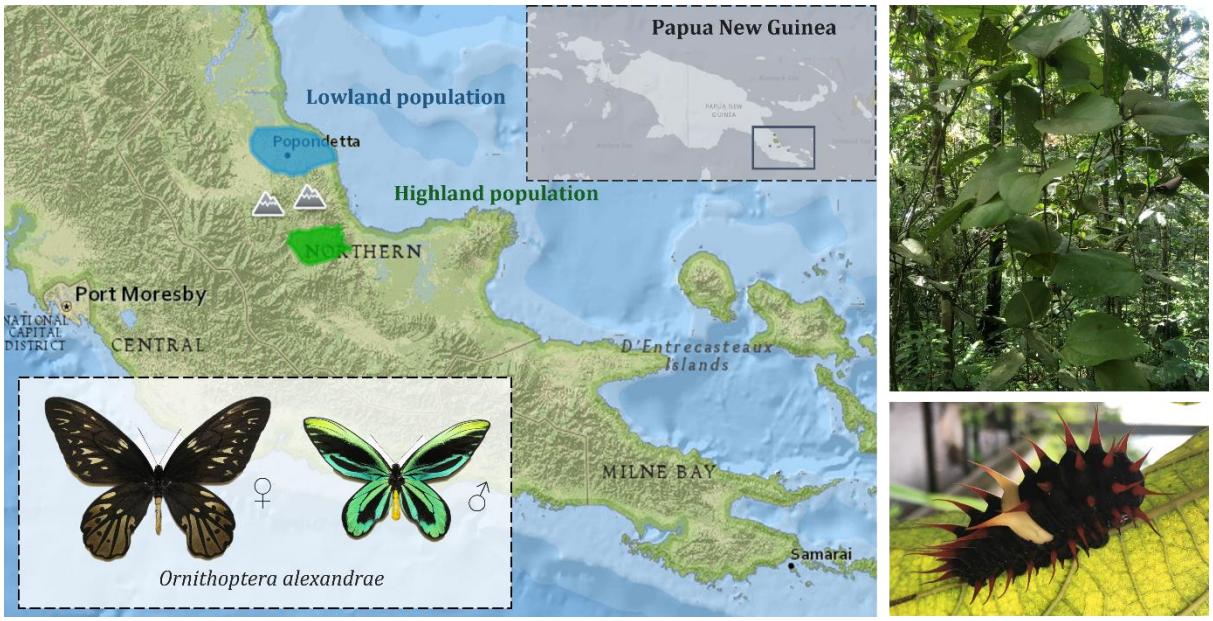

Fig. 1. Left, distribution range of *Ornithoptera alexandrae*, lowland population in blue and highland population in green. Topright, *Aristolochia cf. meridionaliana* plant in its environment. Bottomright, *O. alexandrae* larva. Photos: Fabien L. Condamine. Papua New Guinea gray map is from IUCN Red List https://www.iucnredlist.org/ (Böhm, 2018), and *O. alexandrae* distribution map has been designed on MapMaker https://mapmakerclassic.nationalgeographic.org based on Mitchell et al. (2016).

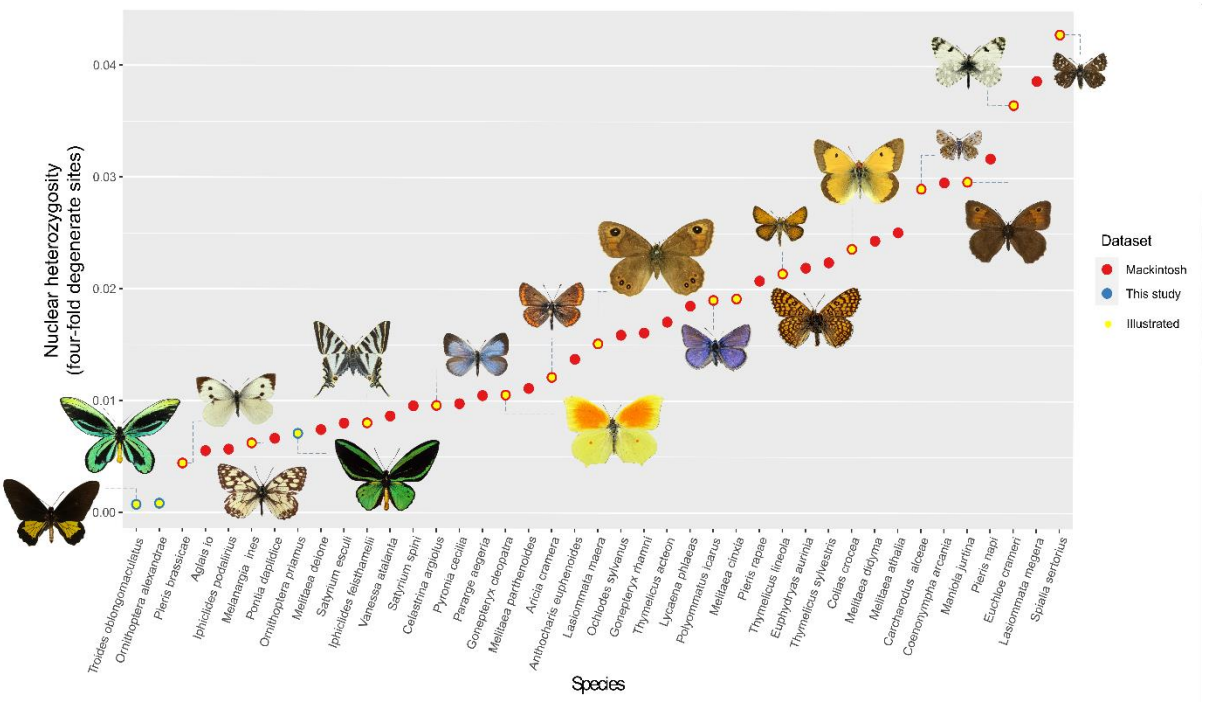

Fig. 2. Level of heterozygosity for butterflies estimated with the four-fold degenerate sites (neutral diversity). Red points are values from Mackintosh et al. (2019) and blue points are from this study. Yellow point indicates illustrated species. Photos (not at scale): *T. oblongomaculatus* (CC BY 4.0, Peter Wing); *O. alexandrae* (Fabien L. Condamine); *P. brassicae* and *G. cleopatra* (CC BY-SA 3.0, Sarefo); *M. ines* (CC BY-SA 4.0, Atylotus); *O. priamus* (Eliette L. Reboud); *I. feisthamelii* and *M. cinxia* (CC BY-SA 4.0, Didier Descouens); *C. argiolus* (CC BY 3.0, Alan Cassidy); *A. cramera* and *P. icarus* (Robin Noel); *L. maera* and *C. crocea* (CC BY-SA 3.0, Vítězslav Maňák); *T. lineola* and *S. sertorius* (CC BY-NC-SA, Peter Huemer); *C. alceae* (CC BY-SA 3.0, Didier Descouens); *M. jurtina* (Public domain, Pekka Malinen); *E. crameri* (Alexander Slutsky).

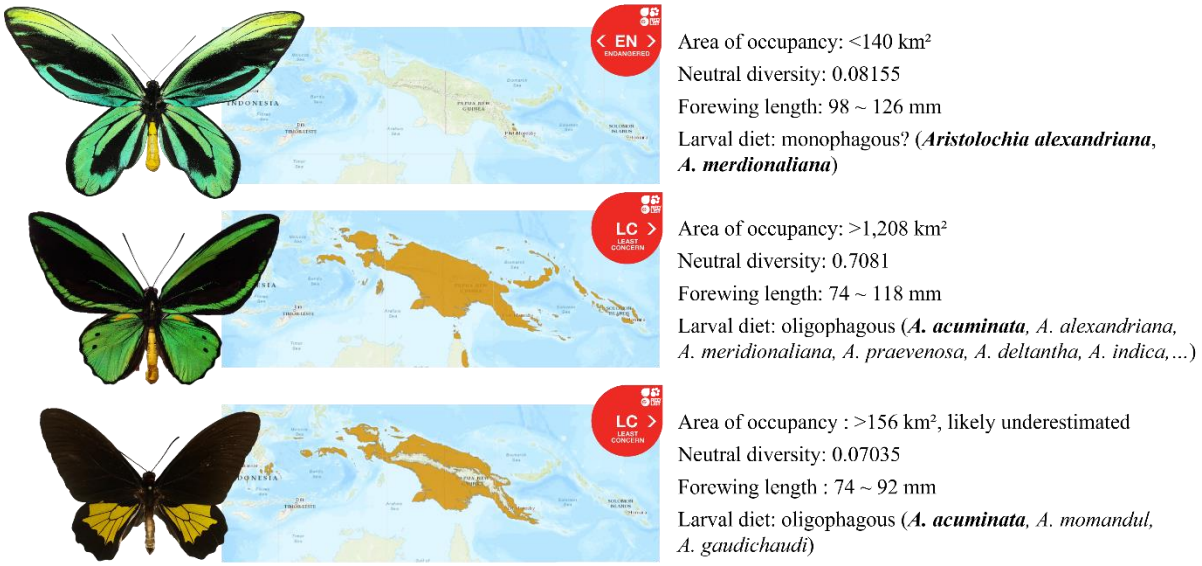

Fig. 3. Comparison of *O. alexandrae*, *O. priamus* and *T. oblongomaculatus*. Distribution maps, IUCN status and areas of occupancy are from IUCN Red List https://www.iucnredlist.org/ (Böhm, 2018). Neutral diversity of *O. alexandrae* is the mean neutral diversity values of the four individuals. Forewing length is from Nakae (2021) and its larval diet is from Mitchell et al. (2016) and Böhm (2018). Larval diets should be taken with caution due to uncertainties of the *Aristolochia* taxonomy. Host plant species in bold font are known to be the main diet of the species. Photos (male specimens, relatively scaled to the mean species forewing value): *O. alexandrae* (Fabien L. Condamine), *O. priamus* (Eliette L. Reboud), *T. oblongomaculatus* (CC BY 4.0, Peter Wing NHM specimen).

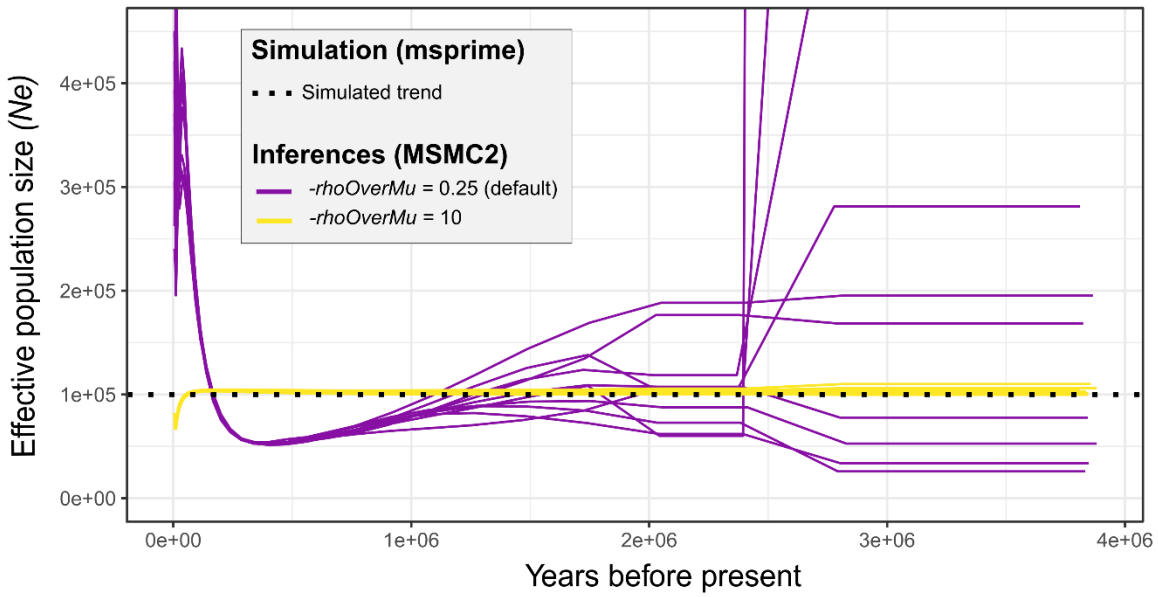

Fig. 4. Performance of MSMC2 inferences on simulated data. Dotted line represents the simulated scenario produced by msprime (Stable *Ne* = 1e5, *r* =1e-8, μ = 1.319e-9). Colored lines represent the demography inferred by MSMC2 on this data with different *-rhoOverMu* settings (ten repetitions each).

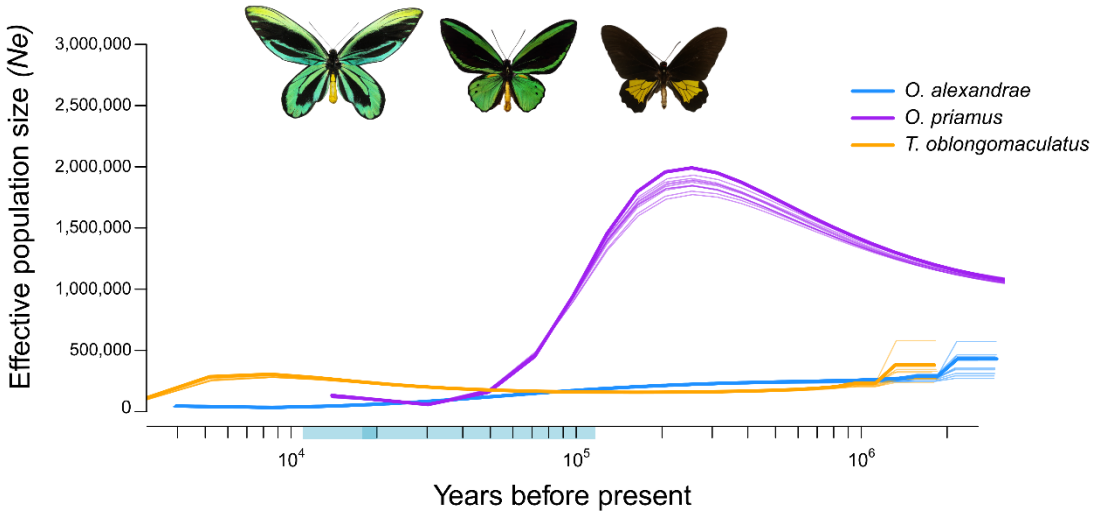

Fig. 5. Estimated historical demography of three birdwing butterfly species. MSMC2 estimates of the effective population size (*Ne*) with *-rhoOverMu* = 10, for *O. alexandrae* (blue), *O. priamus* (purple) and *T. oblongomaculatus* (orange). Bootstraps are represented in clear lines. The pale blue rectangle along the time bar indicates the limits of the last glaciation period (11,700 to 115,000 years ago) with the last glacial maximum in darker blue (19-20,000 years ago). The recent present (last 4,000 years) is not represented.

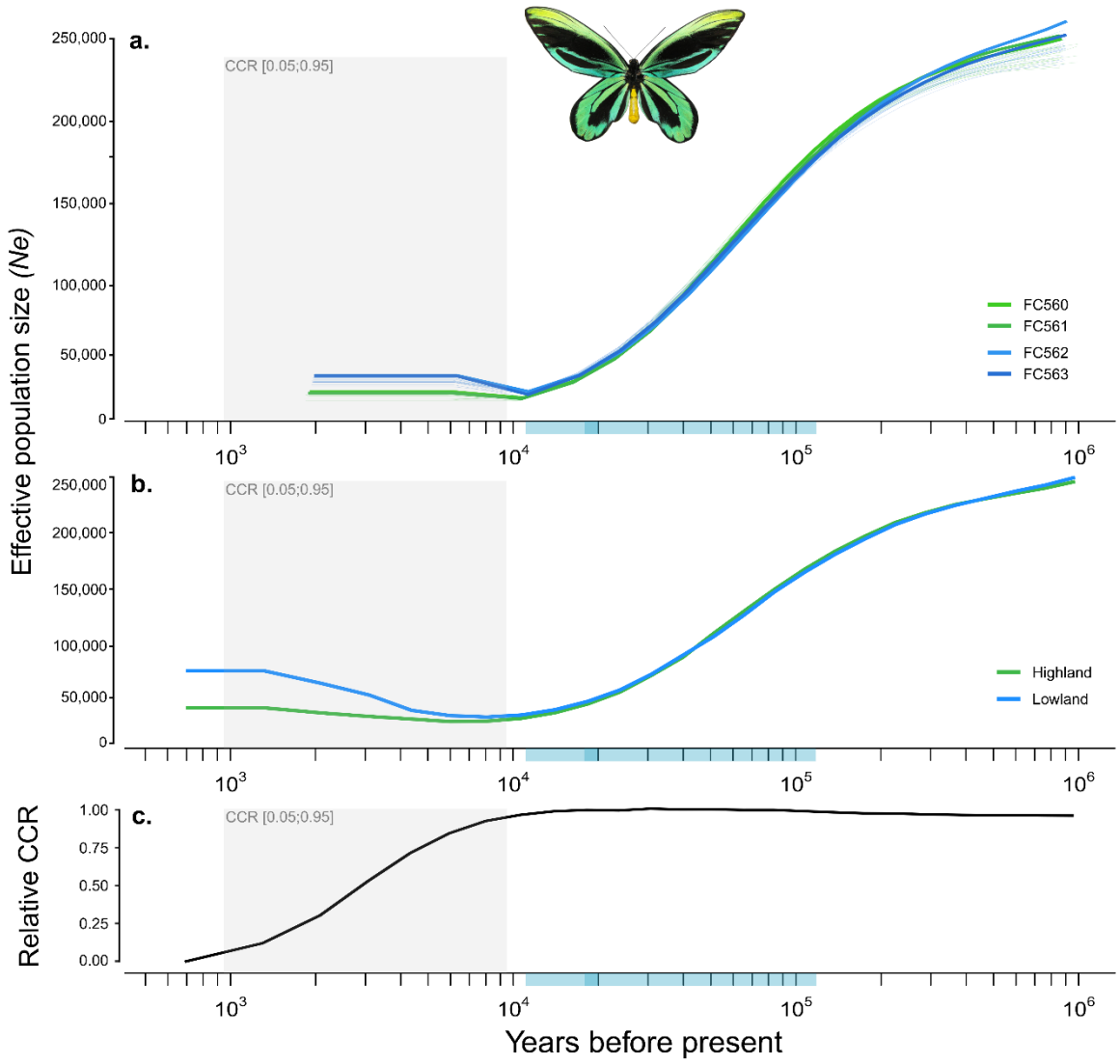

Fig. 6. Demographic inferences of *O. alexandrae* populations. **a.** MSMC estimates for each individual of *O. alexandrae*, with bootstraps represented in clear lines. **b.** MSMC estimates for both populations of *O. alexandrae*. **c.** Relative Cross Coalescent Rate (CCR) estimated with the MSMC analysis, representing the time interval of population divergence. The gray zone represents the 5%-95% CCR ([947; 9460] years ago), and has been reported on each plot. The pale blue rectangle along the time bar indicates the limits of the last glaciation period (11,700 to 115,000 years ago) with the last glacial maximum in darker blue (19-20,000 years ago). For every graph, the recent present (last 700 years) is not represented.

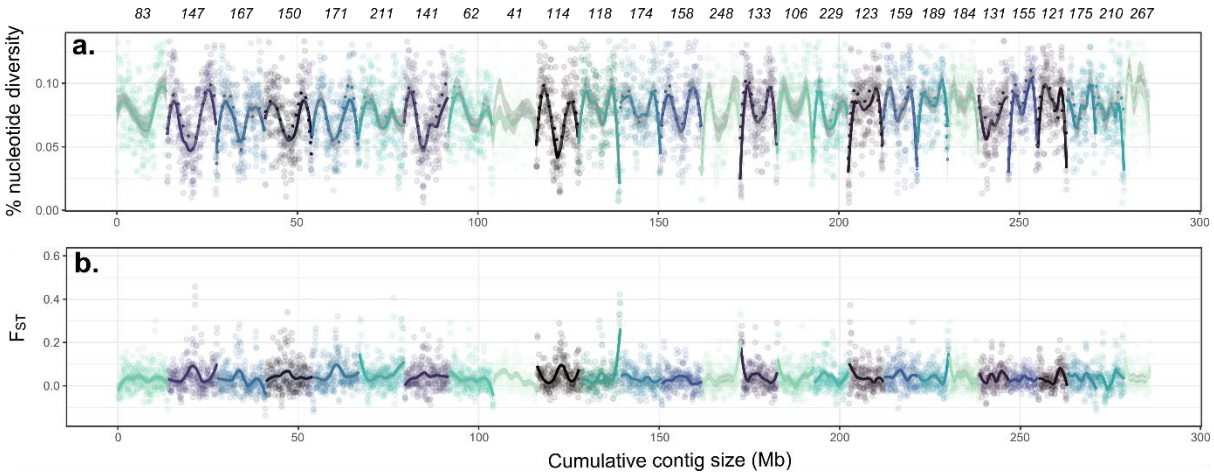

Fig. 7. Genome-wide genetic diversity and differentiation within *Ornithoptera alexandrae*. **a.** Percentage of nucleotide diversity of the genome of *O. alexandrae* for all individuals. **b.** Genetic differentiation (F_{ST}) between the highland and lowland populations. Nucleotide diversity and F_{ST} were computed on non-sex-related contigs >500 kb in length, over sliding windows of 100 kb with an overlap of 50 kb. Alternance of colors indicates the different chromosome-level contigs. Contig names are indicated at the top in italics.

Tables

Table 1. Assembly statistics for the genomes of *O. alexandrae*, *O. priamus* and *T. oblongomaculatus*. LR = long reads, SR = short reads, bp = base pairs. For BUSCO scores, S = single-copy genes, D = duplicated genes, F = fragmented genes, and M = missing genes out of 5,286 genes in odb10 lepidopteran database. Annotation was only done for *O. alexandrae* individuals.

	Ornithoptera alexandrae				*Ornithoptera priamus*	*Troides oblongomaculatus*
	Highland		Lowland			
	FC560 (♂)	FC561 (♀)	FC562 (♀)	FC563 (♀)	FC565 (♂)	FC569 (♀)
Raw data sequenced (Gb) (LR + SR)	30.98 + 33.2	31.82 + 31.0	41.18 + 33.8	29.44 + 37.6	38.99 + 36.9	38.34 + 57.4
Final mean coverage (LR + SR)	66x + 84x	59x + 79x	95x + 70x	70x + 95x	86x + 96x	87x + 137x
Assembly size (bp)	326,746,405	325,591,695	320,565,121	321,134,305	316,500,205	348,219,047
Number of contigs	333	1,222	465	305	862	465
N50 (bp)	10,703,829	7,761,365	9,990,636	11,239,331	4,951,043	5,909,187
Max length (bp)	20,033,826	12,853,001	14,559,367	13,869,660	10,532,357	13,649,974
Nucleotid assembly BUSCO score (%)	S:98.6; D:0.2; F:0.3; M:0.9	S:98.6; D:0.2; F:0.2; M:1.0	S:98.7; D:0.2; F:0.2; M:0.9	S:98.6; D:0.2; F:0.2; M:1.0	S:98.4; D:0.3; F:0.3; M:1.0	S:98.5; D:0.2; F:0.3; M:1.0
Protein coding genes (number and mean size)	17,617 6,095bp	17,449 6,082bp	16,159 6,473bp	16,508 6,325bp	-	-