Intercontinental panmixia despite distinct migration distances in the trans-Saharan butterfly migrant *Vanessa cardui*

Megan S. Reich[1*¶], Daria Shipilina[2*¶], Venkat Talla[2], Farid Bahleman[3], Khadim Kébé[4], Johanna L. Berger[5,6], Niclas Backström[2], Gerard Talavera[6&], Clément P. Bataille[7&]

[1]Department of Biology, University of Ottawa, Ottawa, Ontario, Canada

[2]Evolutionary Biology Program, Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden

[3]SOS Savane - NGO, Tanguiéta, Benin

[4]GRBA-BE, Department of Chemical Engineering and Applied Biology, Polytechnic Higher School of Dakar, Dakar, Senegal

[5]Ecological Networks, Technische Universität Darmstadt, Darmstadt, Germany

[6]Institut Botànic de Barcelona (IBB), CSIC-CMCNB, Barcelona, Catalonia, Spain

[7]Department of Earth and Environmental Sciences, University of Ottawa, Ottawa, Ontario, Canada

*corresponding authors

Email: megan.reich@uottawa.ca (MR); daria.shipilina@ebc.uu.se (DS)

¶These authors contributed equally to this work

&Joint Senior Authors

Abstract

The painted lady butterfly *Vanessa cardui* is renowned for its virtually cosmopolitan distribution and the remarkable long-distance migrations that are part of its annual, multi-generational migratory cycle. In the autumn, *V. cardui* individuals are found north and south of the Sahara, suggesting distinct migratory behaviours within the species. However, the evolutionary and ecological factors shaping these differences in migratory behaviour remain largely unexplored. Here, we performed whole-genome resequencing and analysed the hydrogen and strontium isotopic signatures of 40 *V. cardui* individuals simultaneously collected in the autumn from regions both north and south of the Sahara. Our investigation revealed two main migratory groups: (i) short-distance migrants, journeying from temperate Europe to the circum-Mediterranean region and (ii) long-distance migrants, originating from temperate Europe, crossing the Mediterranean Sea and Sahara, and reaching West Africa, covering a distance of up to over 4,000 km. Despite these stark differences in migration distance, a genome-wide analysis of 813,810 single nucleotide polymorphisms (SNPs) revealed that both short- and long-distance migrants belong to a single intercontinental panmictic population extending from northern Europe to sub-Saharan Africa. Contrary to common biogeographic patterns, the Sahara is not a catalyst for population substructuring in this species. No significant genetic differentiation or signs of adaptation and selection were observed between the two migratory phenotypes (pairwise $F_{ST} = 0.001 \pm 0.006$). Nonetheless, two individuals, belonging to the early arrivals to the Afrotropical region and covering longer migration distances, exhibited some genetic differentiation. The lack of genetic structure between short- and long-distance migrants suggests that migration distance in *V. cardui* is a plastic response to environmental conditions.

Introduction

Migration has evolved several times in insects (1) and is often viewed as an advantageous adaptation, as insects move to favourable environmental conditions, exploit seasonal resources, and escape parasites and predators (2). Many migratory animals show variation in migratory behaviour that results in distinctive spatio-temporal movement patterns. As an innate behaviour in insects, the proximate mechanisms for migration are likely held in the genome (3,4), but the environmental and genetic factors driving intraspecific variation in insect migratory behaviour are only beginning to be understood (5).

The painted lady butterfly *Vanessa cardui* is renowned for its obligate, annual, multi-generational migratory cycles throughout Africa, Asia, Europe, and North America (6,7). As part of a migratory cycle covering between 8 and 10 generations and involving both the Palearctic and Afrotropical regions, late summer and autumn generations migrate southward from temperate Europe towards more suitable breeding grounds. While a substantial portion of these southward migrating butterflies journey across the Mediterranean Sea and Sahara to subsequently breed in sub-Saharan Africa (8–10), others migrate to the circum-Mediterranean region (11,12). The factors driving these differences in migratory behaviour, resulting in the disjunct winter distribution of *V. cardui* across the Sahara, remains unknown.

Migratory phenotypes include observable characteristics such as migratory propensity, timing, orientation, and wing morphology (13). However, migratory behaviours are complex traits and are therefore difficult to quantify, especially in insects. The distance an individual butterfly covers during migration is a consequence of this complex phenotype, which creates a distribution of migration distances across individuals, reflecting the extent of the species' ecological niche. In cases where the suitable niche is geographically continuous, it is expected that individuals at the extremes of the migration distance distribution (i.e., those who migrate the shortest or longest possible distances) are selected against due to the inherent risks associated with extremes. In such scenarios, stabilising selection may favour migrants that cover intermediate distances. Conversely, when the

suitable niche is fragmented by large biogeographic barriers, such as the Sahara, extreme migration distances may be favoured by disruptive selection. Selection for specific migration distances can have important repercussions for survival and growth. For example, pronounced climatic differences between regions north and south of the Sahara influence the composition and quantity of host plant species, predators, parasitoids, and weather events, as well as influencing development time (9). Altogether, this scenario leads us to ask: can we detect the action of adaptive evolutionary forces acting on the extremes of the migration distance distribution and effectively disentangle their action from neutral evolutionary processes?

Differences in migratory routes, distances, and orientation has led to genetic differentiation and subsequent divergent selection in multiple bird species such as the European blackcap *Sylvia atricapilla,* Swainson's thrush *Catharus ustulatus,* and willow warbler *Phylloscopus trochilus* (14–17). Likewise, differences in migration distance could lead to similar spatiotemporal separation of *V. cardui* lineages, resulting in reduced genetic exchange between *V. cardui* that migrate across the Sahara and those that remain in the circum-Mediterranean region, thus leading to genetic differentiation. However, exploring the genetic underpinnings of natural variation in migratory behaviour requires not only whole-genome sequencing data, but also accurate migratory phenotyping (4). Studies on insect migration have lagged behind those of birds and mammals because identifying migratory phenotypes in insects is challenging (18). Extrinsic markers such as biologgers commonly record migratory trajectories for birds and mammals, but few studies have used these techniques on insects (19-21) because insects are generally too small, short-lived, and numerous for these techniques to be applied on a large scale. Instead, studies on insects have mainly relied on observations of physiological state (e.g., wing wear scores), behaviour (e.g., insect-monitoring radar), or regional field monitoring (e.g., spatiotemporal location).

Fortunately, naturally-occurring isotopes are intrinsic markers that have proven increasingly useful for estimating the geographic origin of wild-caught insect migrants and have the potential to identify migratory phenotypes. Isotope geolocation relies on the

observation that insect larvae ingest local environmental isotopic signatures through their diet and incorporate these isotopic signatures into their developing tissues (22,23). Insect wings are relatively inactive tissues and largely preserve the isotopic signature of the natal environment (24,25). Thus, the isotopic signature of the wing of a wild-caught insect can be measured and then compared to a spatial model of isotopic variation (i.e., an isoscape) to estimate the insect's natal origin, a process known as geographic assignment. However, single isotope-based geographic assignment generally yields broad and unspecific regions of natal origin making it challenging to calculate downstream estimates, such as migration distance. More spatially-constrained estimates of natal origin can be obtained by combining multiple isotopes with complementary patterns (26). Dual isotope-based geographic assignment combining hydrogen isotope values (δ^2H), driven by climatic variables (e.g., precipitation; 23,27), and strontium isotope ratios ($^{87}Sr/^{86}Sr$), primarily driven by geological processes (e.g., bedrock type and age; 28), have previously been used to classify migratory behaviours in birds and bats (e.g., 29,30). However, isotopes have not often been leveraged to phenotype migratory behaviour and investigate the potential genomic drivers of variation in migratory behaviour.

Here, we combined isotope-based phenotyping of migration distance with whole-genome resequencing data to investigate the mechanisms driving variation in migratory behaviour for *V. cardui* individuals. By coupling δ^2H and $^{87}Sr/^{86}Sr$ into a dual geographic assignment framework, we estimated the natal origin and migration distance of 40 adult *V. cardui* synchronically collected from north and south of the Sahara in the autumn. A genome-wide population resequencing approach was used to investigate potential genetic structuring associated with contrasting migratory phenotypes. This study endeavours to shed light on the underlying factors shaping the remarkable trans-Saharan migratory journeys of *V. cardui*.

Results

Isotope-based phenotyping reveals differences in migration distance

We estimated the natal origins of 40 *V. cardui* of unknown-origin captured during the autumn southward migration from both north (i.e., Morocco, Spain, Portugal, Malta) and south (i.e., Benin, Senegal) of the Sahara using δ^2H and $^{87}Sr/^{86}Sr$ following the geographic assignment framework of Reich *et al.* (26) (S1 Table; Fig 1D). From the posterior probability surfaces, we then calculated a conservative estimate of migration distance (i.e., minimum distance; S1 Fig; Fig 1C).

Based on migration distance, all individuals were grouped into three categories: (i) long-distance migrants (n = 13, 33%; with minimum migration distances ranging from over 500 to over 4,000 km), (ii) short-distance migrants (n = 3, 7%; with minimum migration distances ranging from over 140 to over 240 km), and (iii) individuals of putatively local origin (n = 24, 60%; having potentially travelled less than 100 km) (Fig 1C). Isotopic signatures corresponding to long-distance migrations were found exclusively in individuals captured south of the Sahara, specifically in Senegal and Benin. Based on the summarised posterior probability surfaces, these individuals seemed to originate either from northwestern Africa or Europe, notably from Portugal, Spain, France, or Sweden (Fig 1B). However, some of the individuals (n = 6, representing 32% of the samples from south of the Sahara; Fig 1F) had a high probability of originating from areas near their capture points (< 100 km) and were designated as “putative locals.” Notably, no individuals from south of the Sahara were classified in the short-distance migration category.

Among the samples collected north of the Sahara, some belonged to the short-distance migration category (n = 3, 14%; Fig 1A). However, the majority of these samples exhibited isotopic signatures consistent with a putatively local origin (n = 18, 86%; Fig 1F). The isotope-based analysis doesn't clarify whether this local signature indicates (i) pre-migratory, locally-sourced offspring of early immigrants to the region or (ii) immigrants from a distant region with an isotopic signature similar to their capture location. We also considered

natural history observations to interpret these patterns. The majority of individuals collected in the circum-Mediterranean region during October and November are likely descendants of earlier arrivals to the collection area. This inference is substantiated by the favourable breeding conditions in the region during late September and October (9,11), as well as the observed presence of eggs, larvae, and recently-emerged adults (i.e., expelling meconium) at the time of collection in Morocco, Malta, and Portugal. Thus, we categorise any samples with a high probability of originating near their capture location as putative locals.

To assess the genomic differentiation between migratory phenotypes, it is crucial to integrate the isotopic signature data with spatiotemporal information from individual samples. Consequently, for subsequent analysis, we grouped individuals that travelled long distances from Senegal and Benin, explicitly excluding individuals with putatively local origins, into the "long-distance" group. To balance sample sizes, we grouped individuals from north of the Sahara, encompassing both the short-distance (n = 3) and putatively local categories (n = 18), into the "short-distance" group, operating under the assumption that these putative locals are the next-generation offspring of short-distance migrants.

Wing morphology and wear are uncorrelated with migration distance

Wing wear scores, a metric scaled from 1 (fresh) to 5 (extremely worn), are often used as a proxy for butterfly age and migratory status (8,31,32). The wing wear scores in our study ranged from 2 (few scales lost) to 5 (extremely worn); samples with wing wear scores of 1 were not selected for our study as they are expected to correspond to recently-emerged, pre-migratory individuals. These wing wear scores showed no correlation with migration distance (Spearman rank correlation, rho = -0.06, p = 0.70).

Butterfly wing size and shape are sometimes correlated with migratory ability (e.g., 33). Here, wing shape was uncorrelated with migration distance (p = 0.2) or the country where the sample was collected (p = 0.1). There was a near-significant relationship between wing shape and wing size (R^2 = 0.30; $F_{1,27}$ = 1.8, p = 0.07). Wing shape differed between the sexes ($F_{1,27}$ = 3.4, p = 0.0002), with the wings of females being more disparate in shape than

those of males (S2 Fig; pairwise comparison, $p = 0.03$). Despite the differences in wing shape between males and females, we did not detect any difference in migration distance by sex (Wilcoxon test; $W = 127$, $p = 0.47$). Additionally, wing shape was influenced by the date of collection, with earlier captures having smaller and more elongated wings (S2 Fig; $F_{1,27} = 3.0$, $p = 0.001$).

Lack of population structure across the Sahara

We explored if the distinct differences in migration distances, coupled with the evident spatiotemporal separation of short- and long-distance migrants, might be reflected in genetic evidence of early separation. This could manifest as population structure driven by neutral processes or as early signs of adaptive divergence among migratory phenotypes. We assessed these patterns at the whole-genome level, employing a comprehensive approach that integrated summary statistics, principal component analysis (PCA), and ancestry proportion inference analysis. Remarkably, all approaches yielded consistent results: a lack of detectable population structure within the dataset and negligible differentiation between groups classified as short- and long-distance migrants.

First, using 13,215,545 high-quality callable sites from both coding and non-coding regions (38 individuals, two removed due to quality issues), we summarised levels of nucleotide diversity (π) within groups and genetic divergence (d_{XY}) between groups of *V. cardui* individuals with distinct migratory behaviour (short-distance migrants: $n = 21$; long-distance migrants: $n = 13$), and calculated the fixation index (F_{st}) as a measure of differentiation (S2 Table). We observed relatively high mean values of nucleotide diversity in both short- and long-distance migrants ($\pi_{autosome} = 0.012$) compared to other Lepidoptera (34). Levels of the pairwise nucleotide divergence ($d_{XYautosome} = 0.012$) between two phenotypes were elevated, reflecting the effect of high levels of diversity within samples and stochasticity of the estimates. The level of genetic differentiation between autosomes was very low, but significant (pairwise $F_{ST} = 0.001 \pm 0.006$, $p < 0.001$). A slightly higher fixation index was observed for the Z-chromosome (pairwise $F_{ST} = 0.009 \pm 0.016$), which is expected

given that the smaller effective population size makes the Z-chromosome more susceptible to the effects of genetic drift. We repeated this analysis using groups based on capture location (i.e., north vs. south of the Sahara), but found no substantial differences between summary statistics (S2 Table). These findings indicate that *V. cardui* that migrate short- versus long-distances have limited genetic differences and share similar allele frequencies, and that the effective population size in general is large.

Secondly, we used PCA to investigate potential genetic structuring. After additional filtering steps, we narrowed down our dataset to 813,810 high-quality single nucleotide polymorphisms (SNPs). This analysis revealed that the majority of samples clustered together with no separation between short- and long-distance migrants (Fig 2A), suggesting an overall absence of population structure. However, within this general pattern, two individuals stood out as outliers in the PCA; one showed differentiation primarily along the first principal component (i.e., 19H128) and the other exhibited differentiation along the second principal component (i.e., 19H115). Both of these outlier samples were collected in September and were among the earliest arrivals to sub-Saharan Africa (9; S1 Table). Moreover, these individuals had undertaken particularly long migratory journeys, with one covering over 2,000 km across the Sahara from northwest Africa or Europe (19H115) and the other over 1,000 km from Portugal, France, Sweden, or Finland (19H128) (S3 Fig). The clear separation of these individuals in the PCA biplot could be indicative of population structuring. We explored the contribution of specific SNPs to the first two principal components with *pcadapt* and identified 924 outlier SNPs that made significant contributions to the observed segregation of individuals. However, it is essential to note that drawing conclusions about the potential processes underlying the observed differences based on data from just two individuals is inherently limited. Nevertheless, we have provided a list of candidate SNPs that may serve as a valuable starting point for future investigations (see Data Availability Statement).

Finally, we tested for population structure in our dataset using two different software implementations to infer individual ancestry proportions (admixture analysis). Specifically, we

employed the sparse non-negative matrix factorization clustering algorithm (SNMF) and fastSTRUCTURE to test models assuming 1 to 5 subpopulations. In both cases, the best model was K = 1, indicating that the data are best described by a single panmictic population (Fig 2B).

No islands of genetic differentiation between short- and long-distance migrants

The assessment of genetic structure using markers distributed across the entire genome may obscure regional differences in genetic differentiation. We therefore used window-based analysis of genetic differentiation between populations (i.e., F_{ST} scans) to quantify genetic differentiation across the genome and search for potential signatures of selection. Our analyses were based on two sample groupings: (i) migratory phenotype (i.e., short- and long-distance migrants) and (ii) capture location (i.e., north and south of the Sahara). Both comparisons revealed uniform differentiation landscapes characterized by a low average pairwise F_{ST}. For autosomal regions, the average F_{ST} was 0.001 and regional peaks had a maximum F_{ST} of 0.113 (Fig 2C; S4 Fig). For the Z-chromosome, the average pairwise F_{ST} was 0.009 with a maximum peak of 0.072. This level of variation is consistent with stochasticity of the normalisation and/or neutral variation, as has previously been demonstrated in simulation studies (35) following simulation regimes applicable to our system. Additionally, none of the local F_{ST} peaks were accompanied by a simultaneous elevation of the genetic divergence and drop in diversity within each population extended over a substantial region, therefore not meeting the expectations for signatures of divergent selection.

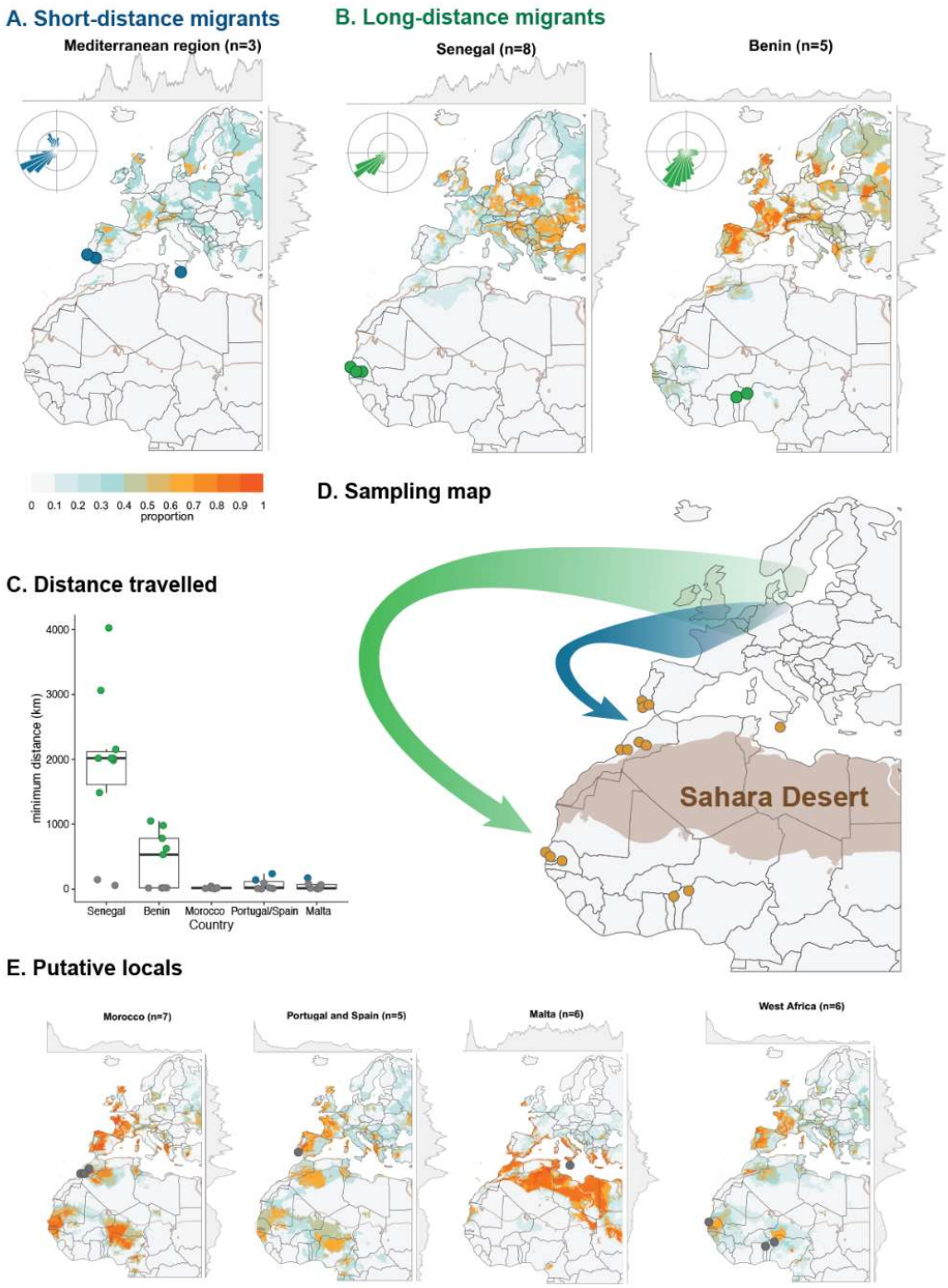

Fig 1. Summarised posterior probability surfaces and subsequent estimates of migration distance (i.e., minimum distance) and direction based on dual δ^2H and $^{87}Sr/^{86}Sr$-based geographic assignment. (A,B,E) Summary maps depict the proportion of individuals with a high probability of natal origin in a given area, as defined by the 2:1 odds ratio. The side margins display the mean proportion by latitude and longitude and only one circle per capture location is shown. The inset rose plot illustrates the average probability-weighted bearing from the estimated natal origin (centre of plot) to the capture locations. Maps summarise the results for **(A)** short-distance migrants, **(B)** long-distance migrants, and **(E)** putative locals captured in the circum-Mediterranean region and West Africa. **(C)** Boxplot displaying the estimated migration distance (km) categorised by the country of capture. Individuals with a minimum distance of less than 100 km were considered putative locals. **(D)** Map indicating the capture locations both north (Morocco, Portugal and Spain, and Malta) and south (Senegal and Benin) of the Sahara. Butterflies captured south of the Sahara generally migrated longer distances (green arrow) compared to those captured north of the Sahara (blue arrow).

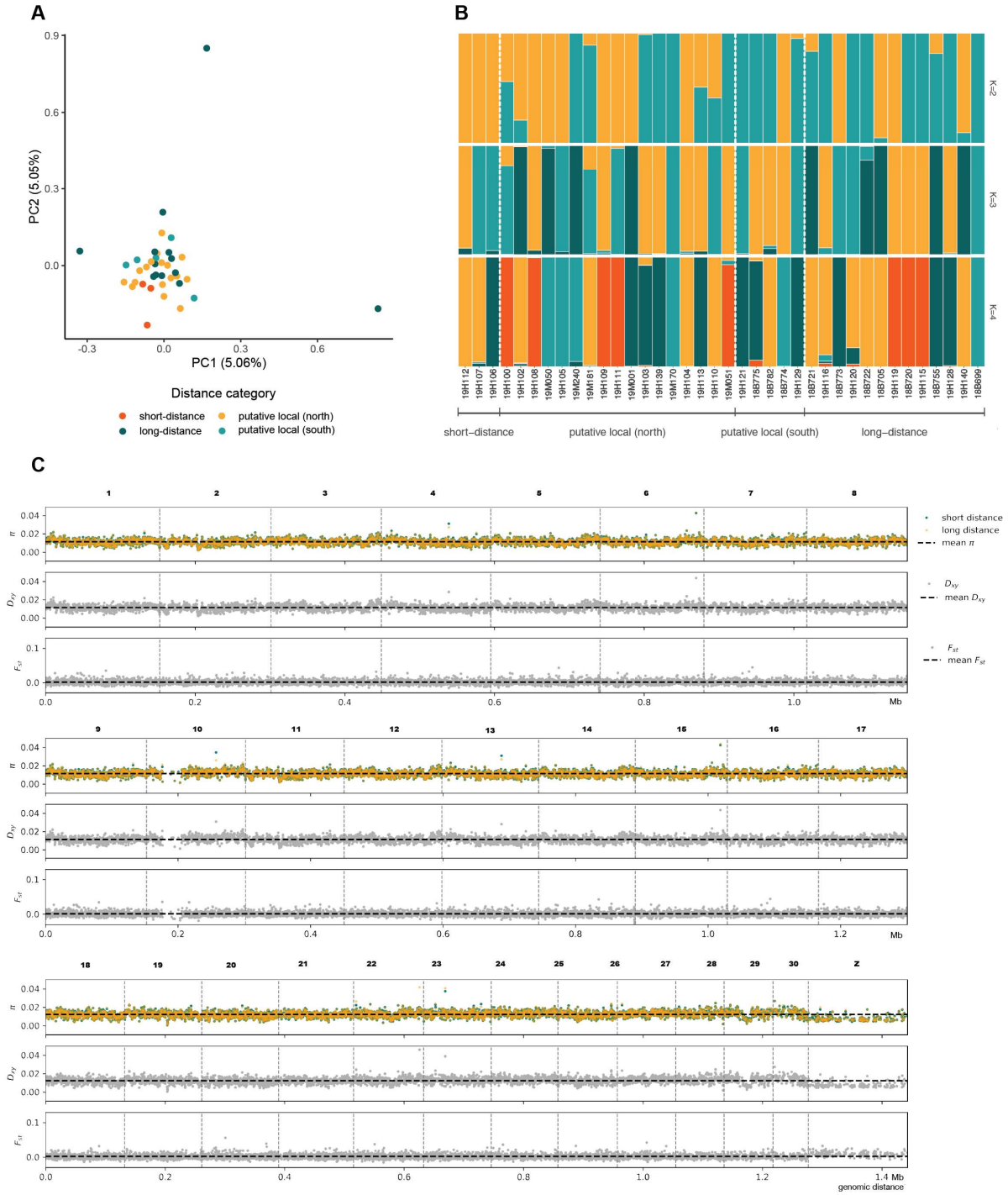

Fig 2. Lack of population structure and genetic differentiation between short- and long-distance migrants. (A) Principal component analysis (PCA) biplot illustrating the genetic variation in the samples; **(B)** Genomic admixture patterns based on assumptions of 2 to 4 ancestral populations (K = 2 to 4). No discernible structure was observed, with the optimal model demonstrating K = 1 (not visually depicted); **(C)** Genome scans showing within-population diversity (π), genetic differentiation (d_{XY}) and fixation index (F_{ST}) between individuals categorised as short- and long-distance migrants. The analysis was performed with sliding, non-overlapping windows of 20 kb. Mean values are indicated by a black dashed line.

Discussion

Contemporary evidence of variation in migratory behaviour

We used isotope-based estimates of migration distance to quantify the natural variation in migration distance exhibited by *V. cardui.* Our findings reveal that butterflies captured in Senegal and Benin during the autumn likely embarked on long migratory journeys, covering up to over 4,000 km as they crossed the Mediterranean Sea and Sahara from Europe. In contrast, *V. cardui* individuals captured in the circum-Mediterranean region in the same season had either undertaken shorter migrations or had isotopic signatures compatible with their capture area (Fig 1). Despite the increasing evidence that *V. cardui* was involved in regular trans-Saharan crossings as part of its annual migratory cycle, this evidence was limited to the monitoring of migratory arrivals in the sub-Sahara (9,10) and δ^2H-based geolocation (8,36). We have expanded this isotope-based evidence with the addition of ^{87}Sr/^{86}Sr, narrowing down the potential geographical sources of these long-distance migrants and confirming that distinctive migration strategies are displayed by *V. cardui* found north and south of the Sahara during the autumn months.

In our study, we found no significant association between estimates of migration distance and wing size, wing shape, or sex. In this regard, *V. cardui* differs from North American monarch butterflies *Danaus plexippus*, which show strong associations between migratory ability and wing morphology in that larger and elongated wings are associated with the increased flight potential of the “super generation” during autumn migration (33,37,38). The variation in migration distance detected in our study appears to have a basis in behaviour rather than wing morphology, which may indicate limited selection on migration distance, as behaviour is thought to be more plastic than morphology (39).

We additionally found no significant association between wing wear score and migration distance. Although wing wear scores have long been used as a proxy for butterfly age and migratory status, with high wing wear scores corresponding to older butterflies that have completed their migratory journeys (e.g., 26,31,40), our results further illustrate that this

proxy is highly unreliable (37). For example, the butterfly with the highest estimated migration distance, up to over 4,000 km (i.e., 18B721), had a low wing wear score of 2 (i.e., “few scales lost”; 32), which is often interpreted as belonging to a recently-emerged adult. This finding is in agreement with Korkmaz et al. (41), who conjecture that high altitude, gliding flight during migration results in less wing damage than low-altitude foraging and breeding behaviour.

Trans-Saharan panmixia

Our genome-wide analysis revealed no evidence of genetic divergence between *V. cardui* individuals migrating to north or south of the Sahara during autumn (Fig 2). This absence of genome-wide differentiation highlights that trans-Saharan migrations are a recurrent, not occasional, phenomenon. Consequently, the *V. cardui* population stretching from Northern Europe to the equatorial regions of sub-Saharan Africa can be characterised as regionally panmictic, confirming the spatial extent of the Afro-Palearctic migratory range. Our findings align well with previous studies using mitochondrial DNA markers (42–44). Given that large biogeographic barriers, such as the Mediterranean Sea and Sahara, do not seem to lead to population substructuring, it is likely that panmixia will be found at even larger scales. However, *V. cardui* migration seems to generally follow latitudinal trajectories, and extending to larger geographical scales might reveal longitudinal structure and population differentiation (e.g., 45). Future studies will need to assess the extent to which global panmixia occurs across the entire distributional range of *V. cardui*, preferably by including genome-wide genetic markers from around the world, including remote islands which sometimes host isolated populations that are phenotypically and genetically differentiated (e.g., 46).

Special attention should be given to understanding how selection and local adaptation act on migratory behaviour in insects, as well as the methods for detecting them. In our study, although we did not detect clear signatures of selection or adaptation, we did detect two outlier samples in the PCA. There could be multiple explanations for the patterns

we detected. First, unlike migratory vertebrates that complete round-trip migrations within a single generation, insect migratory cycles are multi-generational. Multi-generational migratory insects exhibit a complex reticular movement pattern driven by the drastic changes in the location and size of the suitable breeding habitat over the annual migratory cycle (9,11). Therefore, polygenic within-generation selection due to spatially- and temporally-varying selection pressures could lead to allele frequency changes for a portion of the population (47). However, the same alleles selected for in one generation are unlikely to be favoured by selection pressures in the subsequent generation because of spatiotemporal differences in abiotic and biotic conditions during the migratory journey and in the novel breeding habitat, thus diluting the effect of within-generation local selection. Similar frameworks have been proposed for some fish species which show signs of within-generation local selection within panmictic populations (47–49). A similar scenario has been outlined for *D. plexippus*, which experiences selection pressures on wing shape and wing size during long-distance autumn migrations, only to have these pressures released in subsequent generations (37).

Second, it is important to recognize that the exceptionally large effective population size of *V. cardui* (50) and, consequently, weak genetic drift could impede our ability to detect early, or complete, reproductive isolation between short- and long-distance migrants. In these circumstances, many thousands of generations of reproductive isolation would need to elapse before our analysis would be able to detect significant allele frequency shifts. Therefore, it is possible that reproductive isolation has indeed occurred, but remains undetected. Third, it is conceivable that our sampling strategy was unable to detect genetic differentiation between groups of migrants. Notably, the most genetically differentiated migrants (Fig 2B) were on the leading edge of arrivals to West Africa (September 2019) (9) and likely originated in temperate Europe (S3 Fig). These early arrivals may have been exposed to strong environmental cues in temperate Europe, where seasonal changes occur earlier than in regions closer to the equator, prompting the butterflies to initiate their migrations earlier in the season. Early migrants of *D. plexippus* have been shown to have

advantages over their later counterparts, such as enhanced immune defences (51). Consequently, it is possible that *V. cardui* segregates temporally, rather than by distance. However, it is more plausible that this temporal pattern results from within-generation selection, which is likely to be diluted in the subsequent generation. Despite any ephemeral within-generation local selection, we anticipate that the capacity to tolerate heterogeneous environments relies primarily on phenotypic plasticity (52). Ultimately, interpreting our results as indicative of regional panmixia remains the most parsimonious approach.

Panmixia across an Afro-Palearctic distributional range and regular trans-Saharan crossings are exceptional phenomena for insects, as evidenced by the fact that the Palearctic and Afrotropical regions are distinct biogeographic realms with unique faunal compositions, sharing only a relatively small number of species. We found that neither differences in migration distance nor the potential biogeographic barrier led to isolation, selection, or significant allele frequency shifts in any genomic region (Fig 2C). This finding is in contrast to numerous examples in vertebrate species (e.g., 15,17,53,54) and a few instances in insect species (e.g., 55). Some of the species of insects shared between the Palearctic and the Afrotropical regions are suspected to be migratory, particularly among the Lepidoptera, Odonata, and Orthoptera (56,57), and thus it is likely that trans-Saharan gene flow occurs in some of these other species as well. Continental-scale panmixia has been demonstrated in other long-distance migratory insect species, suggesting that panmixia is a prevalent feature among migratory species (but see 58). For example, whole-genome resequencing of *D. plexippus* revealed high gene flow and low genetic differentiation between butterflies located east and west of the Rocky Mountains, despite following geographically separated migration routes (59). Regional panmixia has also been suggested through microsatellite-based population genetics for the green darner dragonfly *Anax junius* in eastern North America (60,61) and global panmixia has been inferred through mitochondrial DNA analysis for the wandering glider dragonfly *Pantala flavescens* (62,63).

The widespread panmixia observed in many migratory insects raises questions about the life history strategies promoting panmixia. First, most insect migrations are multi-

generational, featuring complex movement patterns driven by dynamic changes in breeding habitat, promoting intra-population mixing. Additionally, the tendency of long-distance migratory insects to engage in multiple matings, their lack of courtship behaviour, and their practice of laying numerous eggs (e.g., 60,64,65) hints at an eponymous, yet underexplored, link between panmictic tendencies and mating strategies in migratory insects. Large census population sizes in migratory insects, accompanied by intermittent outbreaks, further drive them towards panmixia. Outbreaks increase the number of successful migrants, potentially eroding local adaptation (66). Outbreaks of *V. cardui* from regions with anomalously high vegetation growth (67) may contribute to the observed lack of genetic structure. Similar population dynamics observed in species like *Libytheana carinenta* (68), *A. junius* (69), and some hoverflies (70) hint at outbreaks being pivotal in shaping the population structure of many migratory insects.

The same life-history factors of migratory insects that appear to promote panmixia among migratory insects (i.e., multi-generational migrations, an absence of assortative mating, and large census population sizes with outbreak dynamics) are also associated with large effective population sizes and high genetic diversity. The multi-generational annual migratory cycle results in a greater number of mutations occurring within the population per unit time, compared to species with longer generation times, and large census population sizes are less susceptible to genetic drift, both of which encourage genetic diversity (71). Indeed, our samples showed relatively high genetic diversity ($\pi_{autosome}$ = 0.012) compared to other butterfly species (34). This observation aligns with the general trend that migratory butterflies tend to possess higher levels of heterozygosity than sedentary species (50). Consequently, this abundant genetic variation provides opportunities for at least some individuals to acclimatise to the heterogeneous environmental conditions that are encountered during their multi-generational migrations, thereby facilitating survival.

Due to the obligate, multi-generational migrations with low philopatry, random mating, and large, outbreaking populations, local adaptation is implausible within the Afro-Palearctic population of *V. cardui*. We suggest that rather than adapting to short or long-distance

migration, a plastic response to the environment determines the differences we observe in migratory phenotype. It is also possible that the differences in migration distance observed here could also be explained, in whole or in part, by extrinsic mechanisms, including differences in parasite load (72), energy storage and metabolism (73), wind-assistance and weather events (74), or even personality (75). However, studies on *D. plexippus* suggest that while migratory state (i.e., resident vs migratory) is genetically determined (76), variation in migration distance is more likely to be due to many loci of small effect or differential gene expression (59), likely controlled by epigenetic mechanisms (4,18,77). Variation in migratory behaviour is likely a response to multiple environmental cues such as resource availability, photoperiod, temperature, humidity, and host plant quality. Recent experiments have demonstrated differences in *V. cardui* gene expression, methylation, and gene regulation due to density of larval conspecifics, larval food availability, and host plant availability for oviposition (78,79), suggesting a strong response to environmental cues. The transcriptional basis of differences in migration distance has been explored in the moth *Helicoverpa armigera* and in *D. plexippus*, yielding promising candidate genes (59,80). Future controlled laboratory studies with *V. cardui* are needed to disentangle the environmental triggers and molecular pathways of phenotypic plasticity in migration distance.

Conclusion

The combination of isotope-based phenotyping of migration distance and comprehensive whole-genome resequencing provided a unique opportunity to investigate phenotype-genotype associations of migratory behaviour in insects, a challenging endeavour until now. We found distinct differences in migration distance, with most individuals captured in Benin and Senegal exhibiting long-distance migrations of up to 4,000 km, transversing the Mediterranean Sea and Sahara from temperate Europe. In contrast, *V. cardui* captured in the circum-Mediterranean region at the same time either undertook short-distance migrations or displayed isotopic signatures indicative of local origins. These differences in migratory behaviour did not have a clear association with wing morphology or wing wear.

Genome-wide comparisons between these migratory groups found negligible genetic differentiation, leading us to conclude that the Afro-Palearctic population of *V. cardui* is panmictic, with regular bidirectional trans-Saharan crossings during the annual migratory cycle. Given the life history traits of *V. cardui*, namely multi-generational migration, their mating strategy, and large population sizes, local adaptation appears implausible within this population. Consequently, future research should explore the role of environmental cues and differential gene expression in shaping the diversity in migration distance.

Materials and Methods

1. Sample collection

Wild, adult *V. cardui* were captured from six countries located both north and south of the Sahara (S1 Table). South of the Sahara, butterflies were captured in Senegal and Benin from August to November of 2018 and 2019 (n = 19). North of the Sahara, butterflies were captured from Morocco, Malta, Portugal, and Spain in November 2019 (n = 21). We used behavioural, demographic and timing evidence to maximise the probability of targeting migratory individuals at these locations. First, we preferentially sampled butterflies with worn wings (i.e., higher wing wear scores) and showing reproductive behaviour (i.e., hilltopping or oviposition) because *V. cardui* typically delay their reproductive behaviour until migration has been completed (81). Second, demographic data in some of the sampling locations supported the arrival of migrants. A monitoring station in Senegal saw a rising number of *V. cardui* adults in the autumn during 2018 and 2019, followed by a peak in larval counts, suggesting that the sudden spike in adults in the autumn is a result of immigration to the region (9). Similarly, a monitoring station in Benin saw high numbers of adults in September and October, followed by high larval counts in October and November (9). Besides this demographic evidence, the time of the year when these butterflies were collected matches the first estimated arrival of migrants at the sampling locations. South of the Sahara, monitoring data does not indicate the presence of *V. cardui* year-round (9). North of the

Sahara, much of the circum-Mediterranean region is unsuitable for *V. cardui* development in July, August, and September (Morocco) due to high temperatures and low host plant availability (40), suggesting that these individuals represent the first arrival of migrants in the region (9, 40).

2. Isotopic analysis

2.1 Hydrogen isotope values

Prior to δ^2H analysis, a forewing from each *V. cardui* was soaked, with agitation, in three successive baths (1 h, 30 min., 10 min.) of 2:1 chloroform:methanol solution to remove surficial dust and lipids, which are known to introduce error into δ^2H measurements (82,83). Samples were cut from the same membranous region of the forewing to reduce intra-individual variation from pigmentation (82) and wing veins (24). Samples were then weighed (0.162 ± 0.009 mg) into silver capsules and surface moisture was dried in a 40 °C oven. Non-exchangeable δ^2H values were determined at the Ján Veizer Stable Isotope Laboratory (University of Ottawa, ON, Canada) using a comparative equilibrium approach (84). All measurements were taken using high temperature (1400 °C) flash pyrolysis (TCEA, Thermo Finnigan, Germany) with a helium carrier passed through a chromium-filled reactor and, after separation, introduced to a Delta V Plus IRMS (Thermo Finnigan, Germany) via a Conflow IV interface (Thermo Finnigan, Germany). A three-point calibration was used to calibrate the sample δ^2H values: CBS (caribou hoof; -157 ± 0.9‰; 85), KHS (kudu horn; -35.3 ± 1.1‰; 85), and USGS43 (human hair; -44.4 ± 2.0‰; 86). Internal standards were measured to assess the quality of the measurements, including one keratin reference standard, USGS42 (human hair; measured: -75.3 ± 0.5‰, n = 4; standard: -72.9 ± 2.2‰ (86)), and two in-house chitin standards, ground and homogenised *Lymantria dispar dispar* (measured: -64.4 ± 1.8‰, n = 6; long-term average: -64 ± 0.8‰) and Alfa Aesar chitin (measured: -22.8 ± 0.7‰, n = 4; long-term average: -22 ± 1.2‰). Based on within-run replicates of the internal standards and repeated sample measurements, precision is estimated to be about ± 2‰. All

reported δ^2H values are normalised to the Vienna Standard Mean Ocean Water - Standard Light Antarctic Precipitation (VSMOW-SLAP) standard scale.

2.2 Strontium isotope ratios

To prepare for $^{87}Sr/^{86}Sr$ analysis, a single forewing from each butterfly was cleaned using pressurised nitrogen gas for 2 min. at 10 psi to remove any surface contaminants (e.g., dust; 26). The wings were then digested in 1 mL HNO_3 (16 M; distilled TraceMetal™ Grade; Fisher Chemical, Canada) using microwave digestion (Anton Paar Multiwave 7000; Austria). The samples were heated from ambient temperature to 250 °C at a steady rate over 20 min. and then left at 250 °C for 15 min. in a pressurised chamber. An aliquot of 200 µL from each sample was separated and diluted to 2 mL of 2% v/v HNO_3 and analysed for Sr content via inductively coupled plasma mass spectrometry (ICP-MS; Agilent 8800 ICP-QQQ, Agilent Technologies Inc., CA, USA) at the Department of Earth and Environmental Science, University of Ottawa (ON, Canada). Calibration standards were prepared using single element certified standards purchased from SCP Science (Montreal, Canada). Samples had an average of 3.1 ± 1.7 ng Sr per mg of wing tissue (n = 40).

The remaining aliquot of sample digest was dried down and re-dissolved in 0.5 mL 6 M HNO_3. The separation of Sr was processed in a 100 µL microcolumn loaded with Sr-spec Resin (100 - 150 µm; Eichrom Technologies, LLC). The matrix was rinsed out using 6 M HNO_3, and Sr was collected with 0.05 M HNO_3. To ensure complete separation, the microcolumn process was repeated. After separation, eluates were dried and re-dissolved in 200 µL 2% v/v HNO_3 for $^{87}Sr/^{86}Sr$ analysis.

The $^{87}Sr/^{86}Sr$ analysis was performed at the Pacific Centre for Isotopic and Geochemical Research (University of British Columbia, BC, Canada) using a Nu-Plasma II high-resolution multi-collector inductively coupled plasma mass spectrometer (MC-ICP-MS; Nu Instruments) coupled to a desolvating nebulizer (Aridus II, CETAC Technologies). The isotopes ^{84}Sr, ^{86}Sr, and ^{87}Sr have isobaric interferences from ^{84}Kr, ^{86}Kr, and ^{87}Rb respectively, and were corrected for using the ^{85}Rb and ^{83}Kr signals. Instrumental mass

fractionation was corrected by normalising $^{86}Sr/^{88}Sr$ to 0.1194 using the exponential law (87). Procedural blanks were negligible. The reproducibility of the $^{87}Sr/^{86}Sr$ measurement for 5 ppb NIST SRM987 is 0.71025 ± 0.00009 (1 SD, n = 138) and for 1.4 ppb NIST SRM987, 0.71019 ± 0.00011 (n = 48). A matrix-matched chitin internal standard, 5 ppb Alfa Aesar chitin, was also used (0.713959 ± 0.00009; n = 3).

2.3 Geographic assignment

Continuous-surface isotope-based geographic assignment was performed via the assignR package (88) in R v4.1.0 (89). The origins of the unknown-origin *V. cardui* were estimated through a probabilistic framework using previously-calibrated isoscapes: (i) a hydrogen isoscape calibrated with residential butterfly samples from across the Palearctic and Afrotropical regions (90), and (ii) a regional bioavailable strontium isoscape (91). The most likely natal origins of each individual was estimated using δ^2H isotopes alone, $^{87}Sr/^{86}Sr$ isotopes alone, and using both isotopes together (Supplementary Material). The 2:1 odds ratio was chosen, in keeping with previous studies (e.g. 31), to create binary surfaces from the posterior probability surfaces of the dual δ^2H and $^{87}Sr/^{86}Sr$-based geographic assignment, representing the top third of the probability distribution as high probability of natal origin (i.e., “1”) and everything else as low probability of natal origin (i.e., “0”). Individual binary surfaces were then summed for each capture location, and divided by the number of individuals, to summarise the results of the geographic assignment into maps of the proportion of individuals with a high probability of natal origin at a given raster cell.

2.4 Distance metric

A consensus has not been reached on the optimal method by which to estimate the distance a migratory individual has travelled from posterior probability surfaces (i.e., the main output of continuous-surface isotope-based geographic assignment) (92). We assessed the ability of seven different metrics to accurately estimate the migration distance of aerial migrants in our study area using synthetic data generated in R v4.1.0 (89). We generated isotopic

signatures for 250 synthetic migrants and then performed dual δ^2H and $^{87}Sr/^{86}Sr$ geographic assignment for each migrant using the pdRaster function in the assignR package (88). Each of the seven metrics was then calculated using the resultant posterior probability surfaces. Most metrics were highly inaccurate, and could under- or over-estimate the actual great circle distance by thousands of kilometres (S1 Fig). However, the “minimum distance” metric consistently underestimated migration distance for our study area, and acts as conservative estimate of migration distance (S1 Fig). Therefore, in this study, we used the minimum distance metric to estimate the distance travelled. This metric measures the shortest distance between the capture location and an isopleth threshold, here defined as the top third of the probability distribution (0.3 isopleth; e.g., 93,94).

Many butterflies had an isotopic signature similar to the local isotopic signature of their capture location. We interpreted individuals with a highly probable natal origin at the capture location (as defined by the 2:1 odds ratio) and/or that had a minimum distance less than 100 km as “putative locals.” These individuals are likely either the locally-sourced offspring of the previous migratory generation or were immigrants from a far-off region with a similar isotopic signature to that of their capture location. Our interpretation of these samples as putative locals is the more parsimonious approach.

3. Geometric morphometric analysis

Differences in migration distance or trajectories could potentially be explained by intrinsic factors such as sex, body size, or wing shape. To test for differences in wing shape, photographs of each sample’s forewing were taken and used for morphometrics analysis. Since we preferentially chose samples that had higher wing wear scores, extensive damage to the wings made landmarking of many samples impossible. Therefore, only 28 of the 40 samples were landmarked using TPSDig2 v.2.32 (95) using 12 conventional landmarks for Nymphalid butterflies (e.g., 96,97). Using the R package geomorph (98), a generalised procrustes analysis removed the translational, rotational, and scaling variation of the shape.

A procrustes regression was used to test the effects on wing shape of sex, month of collection, estimated migration distance, and butterfly wing size.

4. Genomic analysis and genotyping

4.1 Whole genome resequencing

Library preparation for all 40 samples was done using TruSeq PCR-free kits (Illumina, San Diego, USA) at the Institut Botànic de Barcelona (IBB – CSIC, Barcelona, Spain). Whole-genome resequencing was performed with paired-end 2x151 bp reads on one Illumina NovaSeq 6000 S4 lane at the National Genomics Infrastructure (NGI; Stockholm, Sweden) to generate a mean depth of coverage per individual of about 48 reads.

4.2 SNP calling

A modified Sarek Nextflow pipeline (99) was used for preprocessing of the whole genome resequencing data and for detecting single-nucleotide polymorphisms (SNPs). At the first step of the Sarek pipeline, we evaluated the quality of the raw reads with FastQC (version 0.11.9; 100). Reads were stringently trimmed using fastp (101,102) to remove (i) sequences from the 5` end (12 bp) and 3` end (3 bp) of each read, (ii) adaptor sequences, (iii) low quality reads (overall Phred score < 30), and (iv) short reads (< 30 bp). Read quality after trimming was re-evaluated using FastQC. At the next step, reads were mapped to the reference genome (103) using BWA-MEM (version 0.7.17; 104) and then sorted and indexed using SAMtools (version 1.9; 105). A mean of 96.6% of the reads were mapped to the reference genome. Optical and PCR duplicates were marked using GATK MarkDuplicates (version 4.1.4.1; 106). These duplicates, as well as sequences with low mapping quality (MAPQ < 30), were removed using SAMtools. Since prior SNP data for *V. cardui* were unavailable, the base calibration (BQSR) step module in the Sarek pipeline was skipped.

The first steps of variant calling were performed using GATK Best Practices (107), as implemented through the Sarek pipeline. In order to decrease computational time, the Sarek pipeline calls each chromosome separately. MultiQC (version 1.8) was used to assess the

quality of the output from all steps of the pipeline. The resulting per chromosome vcf files included from 1.3 to 3 million unfiltered SNPs (78.9 million SNPs total). The last step of variant calling was performed outside of the pipeline to accommodate invariant site calling (-all-sites option in GenotypeGVCFs), which is necessary for reliable downstream estimation of population genetic summary statistics (such as π and d_{XY}). This variant filtering included multiple steps: (i) hard filtering, (ii) removing SNPs overlapping with TE annotation, (iii) separate quality and coverage filtering for variant and invariant sites, (iv) merging of variant and invariant site vcfs for the F_{ST} estimation. Hard filtering was performed on raw vcf files, which included both variant and invariant sites. Hard filtering is meant to substitute for the variant quality score recalibration (VQSR) step in GATK Best Practices (107). Using bcftools v1.15 (108), we filtered for variant quality by depth (QD < 2.0), Fisher strand bias (FS > 60.0), root mean square mapping quality (MQ < 40.0), u-based z-approximation from the Rank Sum Test for mapping qualities (MQRankSum < -12.5), and u-based z-approximation from the Rank Sum Test for site position (ReadPosRankSum < -8.0). We proceeded to remove indels and multivariate sites (--exclude-types indels, other --max-alleles 2) and reindexed the resulting vcf file. Next, the coordinates of transposable elements and simple repeats were obtained with a custom bash script from a previously published annotation of transposable elements (109), and used to remove SNPs appearing in these regions.

The quality of the resulting files (one vcf file per chromosome) was evaluated using summary statistics implemented with vcftools (i.e., allele frequency (--freq2), mean depth (--depth), mean per site depth (--site-mean-depth), per site quality (--site-quality), per sample missingness (--missing-indv), per site missingness (--missing-indv), heterozygosity estimation (--het); 110). The obtained values were plotted with custom python scripts for further evaluation. Two individuals were removed due to quality issues (Sample IDs: 19H137 and 19H114). Variant and invariant sites were split into two files for downstream analysis. Informed by the summary statistics, we applied filters for quality and depth to the variant sites vcf file (i.e., minor allele frequency (--max-maf 0), mean depth value (--min-meanDP 30), mean depth value (--max-meanDP 80)). We also applied additional filters to the

invariant sites vcf file (i.e., quality value (--minQ 30), mean depth value (--min-meanDP 30)). Separate files with both variant and invariant calls were kept as separate chromosomes to decrease computational time. The W chromosome was also removed from the analysis. A region with divergent values of populational statistics was detected in the first third of chromosome 10 corresponding to a large area of transposable element proliferation (Fig 2C; 109). Since this region likely does not carry information about genetic structure it was also filtered from the population structure analyses. On a technical note, it is important to mention that this region was retained after stringent SNP filtration, and future research on *V. cardui* should be aware that this region may necessitate special treatment (e.g., manual filtering).

4.3 Population genetic analysis

For PCA and admixture analysis, variant files per chromosome were concatenated. To investigate genetic variation between individuals, a PCA was performed using PLINK v1.90 b 4.9 (Fig 2C). For this analysis, we additionally filtered out singleton SNP sites (i.e., MAF = 0.03), and SNPs that were in linkage (i.e., SNPs with r^2 values > 0.1 using sliding windows of 50 SNPs with a step size of 10 SNPs); a total of 813,810 high quality variants remained after filtering. Outliers in the first two principal components were detected using the R package *pcadapt* (111) using Bonferroni-corrected p-values ($\alpha = 0.05$). The PCA was repeated excluding the two outliers, but no additional structure was detected (S5 Fig).

Reverse migrations from sub-Saharan Africa to the Maghreb are thought to occur in the autumn, even while southwards trans-Saharan migrations are still underway (8). All of the samples captured in Morocco were designated as putative locals, but the posterior probability surfaces of some showed a possibility of originating south of the Sahara (Fig 1E). To test whether these samples could have biased our results, we repeated the PCA without samples collected in Morocco (S5 Fig). The resulting PCA was based on 647,906 SNPs and showed a single cluster with the same outlier individuals as the original PCA.

After excluding SNPs on the Z chromosome and in coding sequences, genetic structure was also assessed with a likelihood method, the sparse non-negative matrix

factorization clustering algorithm, SNMF, in the R package *LEA* v3.6.0 (112). Each number of population clusters (K) from 1 to 5 was run independently with 5 repetitions with the regularisation parameter set to 100, after which cross-validation errors calculated using the cross-entropy criterion were compared. This process was repeated using a Bayesian framework via fastSTRUCTURE (113) using the simple allele frequency prior with 5 repetitions. The number of population clusters from 1 to 5 that best explained the observed structure was chosen with the chooseK.py tool.

4.4 Genomic scan for selection

To detect any fine-scale regions of differentiation along the genome, a genome scan was performed on groupings based on capture location (i.e., north vs. south of the Sahara) and migratory phenotype (i.e., long vs. short migration distance). For the grouping based on migratory phenotype, individuals from south of the Sahara that were putative locals (n = 6) were removed, while putative locals from north of the Sahara (n = 19) were included with the short-distance migrants (n = 3) to balance the sample size. For genome scans we retained SNPs falling into coding regions and having minor allele frequencies. We used pixy v1.2.7 (114,115) to calculate nucleotide diversity (π), absolute divergence (d_{XY}), and genetic differentiation (F_{ST}; Hudson method (-hudson)) in 20 kb non-overlapping windows using all sites. Windows with less than 10% of sites represented (i.e., < 2000 bp) were removed from the analysis. Resulting per window statistics were summarized using custom python scripts. For low levels of F_{ST}, we tested if the index significantly differed from 0 using a t-test in the python scipy package (p-value: 1.3e-216).

Author Contributions

GT and CB conceived the idea for the study and organised the sample collection. GT, MR, KK, and FB collected the samples. MR and CB performed the isotope analysis component. DS, MR, GT, VT, and NB performed the genomic analysis component. JB did the landmarking for the geometrics morphometrics analysis. MR and DS led the analysis and

writing of the manuscript. All authors contributed to the drafts and gave final approval for publication.

Acknowledgements

The authors acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) and the Swedish National Infrastructure for Computing (SNIC) in Uppsala, partially funded by the Swedish Research Council through grant agreement no. 2022-06725 and no. 2018-05973. We are grateful to members of the Backström lab and Sean Stankowski for insightful discussions about the genome scan analysis. We would like to thank Smita Mohanty for help with the ICP-MS, Kerry Klassen and Paul Middlestead at the Ján Veizer Stable Isotope Laboratory for help with the hydrogen isotope analysis, and Kathy Gordon at the Pacific Centre for Isotopic and Geochemical Research for help with the strontium isotope analysis. We would like to thank Leonardo Dapporto, Leonardo Platania, and Aurora García-Berro for sharing code and knowledge related to the wing geometric morphometrics analysis, and Eude Goudégnon, Roger Vila, Mattia Menchetti, and Tomasz Suchan for help in sample collection. We would also like to thank Andrea Contina, Gabriel Bowen, and Michael Wunder for sharing code and insight relating to the migration distance estimates.

Funding

This study was funded by grant 2018-00738 of the New Frontiers in Research Fund (Government of Canada) to GT and CB; by grants PID2020-117739GA-I00 MCIN / AEI / 10.13039/501100011033 and WW1-300R-18 (National Geographic Society) to GT; and by grant LINKA20399 from the CSIC iLink program to NB, CB and GT. MR was supported by

the Queen Elizabeth II Graduate Scholarship in Science and Technology (QEII-GSST), the Ontario Graduate Scholarship, and the 2021 ORIGIN project (NSF award DBI-1565128). NB is supported by a research grant from the Swedish Research Council FORMAS (grant 2019-00670). DS is supported by the Thurberg postdoctorate fellowship program of Swedish Collegium for Advanced Science (Knut and Alice Wallenberg Foundation).

Data Availability Statement

The raw sequence data is deposited in the European Nucleotide Archive (ENA). All data and code pertaining to the isotopic, geometric morphometric, and genomic analyses, including bash, R, and python scripts, can be found on the Open Science Foundation.

References

1. Dingle H, Drake VA. What is migration? BioScience. 2007;57(2):113–21. doi: 10.1641/B570206
2. Chapman JW, Reynolds DR, Wilson K. Long-range seasonal migration in insects: Mechanisms, evolutionary drivers and ecological consequences. Ecology Letters. 2015;18(3):287–302. doi: 10.1111/ele.12407
3. Dingle H. Evolutionary genetics of migration. In: Migration: The biology of life on the move. Oxford University Press; 2014. p. 248–68.
4. Merlin C, Liedvogel M. The genetics and epigenetics of animal migration and orientation: Birds, butterflies and beyond. J Exp Biol. 2019;222:1–12. doi: 10.1242/jeb.191890
5. Shaw AK. Causes and consequences of individual variation in animal movement. Mov Ecol. 2020;8(1):1–12. doi: 10.1186/s40462-020-0197-x
6. Shields O. World distribution of the *Vanessa cardui* group (Nymphalidae). J Lepidopterists Soc. 1992;46(3):235–8.
7. Williams CB. The migrations of the painted lady butterfly, *Vanessa cardui* (Nymphalidae), with special reference to North America. Lepidopterists Soc.

1970;24(3).

8. Stefanescu C, Soto DX, Talavera G, Vila R, Hobson KA. Long-distance autumn migration across the Sahara by painted lady butterflies: Exploiting resource pulses in the tropical savannah. Biol Lett. 2016;12:1–4. doi: 10.1098/rsbl.2016.0561

9. Talavera G, García-Berro A, Talla VNK, Ng'iru I, Bahleman F, Kébé K, et al. The Afrotropical breeding grounds of the Palearctic-African migratory painted lady butterflies (*Vanessa cardui*). Proc Natl Acad Sci. 2023;120(16):e2218280120. doi: 10.1073/pnas.2218280120

10. Talavera G, Vila R. Discovery of mass migration and breeding of the painted lady butterfly *Vanessa cardui* in the Sub-Sahara: The Europe-Africa migration revisited. Biol J Linn Soc. 2016;120(2):274–85. doi: 10.1111/bij.12873

11. Menchetti M, Guéguen M, Talavera G. Spatio-temporal ecological niche modelling of multigenerational insect migrations. Proc R Soc B Biol Sci. 2019;286(1910):20191583. doi: 10.1098/rspb.2019.1583

12. Stefanescu C, Páramo F, Åkesson S, Alarcón M, Ávila A, Brereton T, et al. Multi-generational long-distance migration of insects: Studying the painted lady butterfly in the Western Palaearctic. Ecography. 2013;36(4):474–86. doi: 10.1111/j.1600-0587.2012.07738.x

13. Turbek SP, Scordato ESC, Safran RJ. The role of seasonal migration in population divergence and reproductive isolation. Trends Ecol Evol. 2018;33(3):164–75. doi: 10.1016/j.tree.2017.11.008

14. Delmore K, Fox JW, Irwin DE. Dramatic intraspecific differences in migratory routes, stopover sites and wintering areas, revealed using light-level geolocators. Proc R Soc B Biol Sci. 2012;279(1747):4582–9. doi: 10.1098/rspb.2012.1229

15. Delmore K, Illera JC, Pérez-Tris J, Segelbacher G, Lugo Ramos JS, Durieux G, et al. The evolutionary history and genomics of European blackcap migration. eLife. 2020;9:e54462. doi: 10.7554/eLife.54462

16. Mueller JC, Pulido F, Kempenaers B. Identification of a gene associated with avian

migratory behaviour. Proc R Soc B Biol Sci. 2011;278(1719):2848–56. doi: 10.1098/rspb.201

17. Sokolovskis K, Lundberg M, Åkesson S, Willemoes M, Zhao T, Caballero-Lopez V, et al. Migration direction in a songbird explained by two loci. Nat Commun. 2023;14(1):165. doi: 10.1038/s41467-023-35788-7
18. Liedvogel M, Åkesson S, Bensch S. The genetics of migration on the move. Trends Ecol Evol. 2011;26(11):561–9. doi: 10.1016/j.tree.2011.07.009
19. Knight SM, Pitman GM, Flockhart DTT, Norris DR. Radio-tracking reveals how wind and temperature influence the pace of daytime insect migration. Biol Lett. 2019;15(7): 20190327. doi: 10.1098/rsbl.2019.0327
20. Menz MHM, Scacco M, Bürki-Spycher HM, Williams HJ, Reynolds DR, Chapman JW, et al. Individual tracking reveals long-distance flight-path control in a nocturnally migrating moth. Science. 2022;377(6607):764–8. doi: 10.1126/science.abn1663
21. Wikelski M, Moskowitz D, Adelman JS, Cochran J, Wilcove DS, May ML. Simple rules guide dragonfly migration. Biol Lett. 2006;2:325–9. doi: 10.1098/rsbl.2006.0487
22. Flockhart DTT, Kyser TK, Chipley D, Miller NG, Norris DR. Experimental evidence shows no fractionation of strontium isotopes ($^{87}Sr/^{86}Sr$) among soil, plants, and herbivores: Implications for tracking wildlife and forensic science. Isotopes Environ Health Stud. 2015;51(3):372–81. doi: 10.1080/10256016.2015.1021345
23. Hobson KA, Wassenaar LI, Taylor OR. Stable isotopes (δD and $\delta^{13}C$) are geographic indicators of natal origins of monarch butterflies in eastern North America. Oecologia. 1999;120(3):397–404.
24. Lindroos EE, Bataille CP, Holder PW, Talavera G, Reich MS. Temporal stability of $\delta^{2}H$ in insect tissues: Implications for isotope-based geographic assignments. Front Ecol Evol. 2023;11:1060836. doi: 10.3389/fevo.2023.1060836
25. Reich MS, Kindra M, Dargent F, Hu L, Flockhart DTT, Norris DR, et al. Metals and metal isotopes incorporation in insect wings: Implications for geolocation and pollution exposure. Front Ecol Evol. 2023;11:1085903. doi: 10.3389/fevo.2023.1085903

26. Reich MS, Flockhart DTT, Norris DR, Hu L, Bataille CP. Continuous-surface geographic assignment of migratory animals using strontium isotopes: A case study with monarch butterflies. Methods Ecol Evol. 2021;00:1–13. doi: 10.1111/2041-210X.13707
27. Wassenaar LI, Hobson KA. Natal origins of migratory monarch butterflies at wintering colonies in Mexico: New isotopic evidence. Proc Natl Acad Sci U S Am. 1998;95:15436–9.
28. Bataille CP, Crowley BE, Wooller MJ, Bowen GJ. Advances in global bioavailable strontium isoscapes. Palaeogeogr Palaeoclimatol Palaeoecol. 2020;555:109849. doi: 10.1016/j.palaeo.2020.109849
29. Crowley BE, Bataille CP, Haak BA, Sommer KM. Identifying nesting grounds for juvenile migratory birds with dual isotope: An initial test using North American raptors. Ecosphere. 2021;12(10):0–21. doi: https://doi.org/10.1002/ecs2.3765
30. Kruszynski C, Bailey LD, Courtiol A, Bach L, Bach P, Göttsche M, et al. Identifying migratory pathways of Nathusius' pipistrelles (*Pipistrellus nathusii*) using stable hydrogen and strontium isotopes. Rapid Commun Mass Spectrom. 2021;35(6):1–11. doi: 10.1002/rcm.9031
31. Flockhart DTT, Wassenaar LI, Martin TG, Hobson KA, Wunder MB, Norris DR. Tracking multi-generational colonization of the breeding grounds by monarch butterflies in eastern North America. Proc R Soc B. 2013;280(20131087). doi: 10.1098/rspb.2013.1087
32. Miller NG, Wassenaar LI, Hobson KA, Norris R. Migratory connectivity of the monarch butterfly (*Danaus plexippus*): Patterns of spring re-colonization in eastern North America. PLoS ONE. 2012;7(3):e31891. doi: 10.1371/journal.pone.0031891
33. Altizer S, Davis A. Populations of monarch butterflies with different migratory behaviors show divergence in wing morphology. Evolution. 2010;64(4):1018–28. doi: 10.1111/j.1558-5646.20
34. Mackintosh A, Laetsch DR, Hayward A, Charlesworth B, Waterfall M, Vila R, et al. The

determinants of genetic diversity in butterflies. Nat Commun. 2019;10(1):3466. doi: 10.1038/s41467-019-11308-4

35. Lohse K. Come on feel the noise – from metaphors to null models. J Evol Biol. 2017;30(8):1506–8. doi: 10.1111/jeb.13109
36. Talavera G, Bataille C, Benyamini D, Gascoigne-Pees M, Vila R. Round-trip across the Sahara: Afrotropical painted lady butterflies recolonize the Mediterranean in early spring. Biol Lett. 2018;14:20180274. doi: 10.1098/rsbl.2018.0274
37. Flockhart DTT, Fitz-gerald B, Brower LP, Derbyshire R, Altizer S, Hobson KA, et al. Migration distance as a selective episode for wing morphology in a migratory insect. Mov Ecol. 2017;5(7):1–9. doi: 10.1186/s40462-017-0098-9
38. Li Y, Pierce AA, de Roode JC. Variation in forewing size linked to migratory status in monarch butterflies. Anim Migr. 2016;3:27–34. doi: 10.1515/ami-2016-0003
39. West-Eberhard MJ. Phenotypic plasticity and the origins of diversity. Annu Rev Ecol Syst. 1989;20(1):249–78.
40. Stefanescu C, Puig-Montserrat X, Samraoui B, Izquierdo R, Ubach A, Arrizabalaga A. Back to Africa: Autumn migration of the painted lady butterfly *Vanessa cardui* is timed to coincide with an increase in resource availability. Ecol Entomol. 2017;42(6):737–47. doi: 10.1111/een.12442
41. Korkmaz R, Rajabi H, Eshghi S, Gorb SN, Büscher TH. The frequency of wing damage in a migrating butterfly. Insect Sci. 2022;0:1-11. doi: 10.1111/1744-7917.13153
42. Dapporto L, Cini A, Vodă R, Dincă V, Wiemers M, Menchetti M, et al. Integrating three comprehensive data sets shows that mitochondrial DNA variation is linked to species traits and paleogeographic events in European butterflies. Mol Ecol Resour. 2019;19(6):1623–36. doi: 10.1111/1755-0998.13059
43. Pfeiler E, Markow TA. Population connectivity and genetic diversity in long-distance migrating insects: Divergent patterns in representative butterflies and dragonflies. Biol J Linn Soc. 2017;122:479–86.
44. Vodă R, Dapporto L, Dincă V, Shreeve TG, Khaldi M, Barech G, et al. Historical and

contemporary factors generate unique butterfly communities on islands. Sci Rep. 2016;6(28828):1-11. doi: 10.1038/srep28828

45. Singh VK, Kumar S, Awasthi S, Gupta SK. Phylogeography and population expansion of Western Himalayan highly-erratically migratory painted lady butterfly (Lepidoptera: Nymphalidae: *Vanessa cardui*). J Asia-Pac Entomol. 2021;24(4):1116–21. doi: 10.1016/j.aspen.2021.10.005
46. Alvial IE, Vargas HA, Marinov M, Esquivel C, Araya J, Araya-Donoso R, et al. Isolation on a remote island: Genetic and morphological differentiation of a cosmopolitan odonate. Heredity. 2019;122(6):893–905. doi: 10.1038/s41437-018-0165-z
47. Ehrlich MA, Wagner DN, Oleksiak MF, Crawford DL. Polygenic selection within a single generation leads to subtle divergence among ecological niches. Genome Biol Evol. 2021;13(2). doi: 10.1093/gbe/evaa257
48. Als TD, Hansen MM, Maes GE, Castonguay M, Riemann L, Aarestrup K, et al. All roads lead to home: Panmixia of European eel in the Sargasso Sea. Mol Ecol. 2011;20(7):1333–46. doi: 10.1111/j.1365-294X.2011.05011.x
49. Pujolar JM, Jacobsen MW, Als TD, Frydenberg J, Munch K, Jónsson B, et al. Genome-wide single-generation signatures of local selection in the panmictic European eel. Mol Ecol. 2014;23(10):2514–28. doi: 10.1111/mec.12753
50. García-Berro A, Talla V, Vila R, Wai HK, Shipilina D, Chan KG, et al. Migratory behaviour is positively associated with genetic diversity in butterflies. Mol Ecol. 2023;32(3):560–74. doi: 10.1111/mec.16770
51. Satterfield DA, Wright AE, Altizer S. Lipid reserves and immune defense in healthy and diseased migrating monarchs *Danaus plexippus*. Curr Zool. 2013;59(3):393–402. doi: 10.1093/czoolo/59.3.393
52. Enbody ED, Pettersson ME, Sprehn CG, Palm S, Wickström H, Andersson L. Ecological adaptation in European eels is based on phenotypic plasticity. Proc Natl Acad Sci. 2021;118(4):e2022620118. doi: 10.1073/pnas.2022620118
53. Gu Z, Pan S, Lin Z, Hu L, Dai X, Chang J, et al. Climate-driven flyway changes and

memory-based long-distance migration. Nature. 2021;591:259–564. doi: 10.1038/s41586-021-03265-0

54. Thompson NF, Anderson EC, Clemento AJ, Campbell MA, Pearse DE, Hearsey JW, et al. A complex phenotype in salmon controlled by a simple change in migratory timing. Science. 2020;370(6516):609–13. doi: 10.1126/science.aba9059

55. Niitepõld K, Smith AD, Osborne JL, Reynolds DR, Carreck L, Martin AP, et al. Flight metabolic rate and *Pgi* genotype influence butterfly dispersal rate in the field. Ecology. 2009;90(8):2223–32.

56. Chelli A, Moulaï R. Ecological characterization of the odonatofauna in lotic and lentic waters of northeast Algeria. Ann Société Entomol Fr NS. 2019;55(5):430–45. doi: 10.1080/00379271.2019.1660215

57. Pedgley DE, Reynolds DR, Tatchell GM. Long-range insect migration in relation to climate and weather: Africa and Europe. In: Insect migration: tracking resources through space and time. Cambridge University Press; 1995. p. 3–29.

58. Zhang L, Li Z, Peng Y, Liang X, Wilson K, Chipabika G, et al. Global genomic signature reveals the evolution of fall armyworm in the Eastern hemisphere. Mol Ecol. 2023;00:1-16. doi: 10.1111/mec.17117

59. Talla V, Pierce AA, Adams KL, de Man TJB, Nallu S, Villablanca FX, et al. Genomic evidence for gene flow between monarchs with divergent migratory phenotypes and flight performance. Mol Ecol. 2020;29(14):2567–82. doi: 10.1111/mec.15508

60. Matthews JH. Research in motion: Patterns of large-scale migration in dragonflies and birds [dissertation]. The University of Texas at Austin; 2007.

61. May ML, Matthews JH. Migration in Odonata: A case study of *Anax junius*. In: Dragonflies and damselflies: Model organisms for ecological and evolutionary research. Oxford University Press; 2008. p. 63–78.

62. Troast D, Suhling F, Jinguji H, Sahlén G, Ware J. A global population genetic study of *Pantala flavescens*. PLOS ONE. 2016;11(3):e0148949. doi: 10.1371/journal.pone.0148949

63. Ware J, Kohli MK, Mendoza CM, Troast D, Jinguji H, Hobson KA, et al. Evidence for widespread gene flow and migration in the Globe Skimmer dragonfly *Pantala flavescens*. Int J Odonatol. 2022;25:43–55. doi: 10.48156/1388.2022.1917166

64. Pliske TE. Courtship behavior of the monarch butterfly, *Danaus plexippus* L. Ann Entomol Soc Am. 1975;68(1):143–51. doi: 10.1093/aesa/68.1.143

65. Wiklund C, Friberg M. Testing the migration syndrome: Comparative fecundity of migratory and non-migratory nymphaline butterflies. Ecol Entomol. 2022;47:1061–7. doi: 10.1111/een.13195

66. Chapuis MP, Loiseau A, Michalakis Y, Lecoq M, Franc A, Estoup A. Outbreaks, gene flow and effective population size in the migratory locust, *Locusta migratoria*: A regional-scale comparative survey. Mol Ecol. 2009;18(5):792–800. doi: 10.1111/j.1365-294X.2008.04072.x

67. López-Mañas R, Pascual-Díaz JP, García-Berro A, Bahleman F, Reich MS, Pokorny L, et al. Erratic spatiotemporal vegetation growth anomalies drive population outbreaks in a trans-Saharan insect migrant. Proc Natl Acad Sci. 2022;119(19):3–5. doi: 10.1073/pnas.2121249119

68. Neck RW. Causal analysis of a migration of the snout butterfly*, Libytheana bachmanii larvata* (Strecker) (Liibytheidae). J Lepidopterists Soc. 1983;37(2):121–8.

69. Russell RW, May ML, Soltesz KL, Fitzpatrick JW. Massive swarm migrations of dragonflies (Odonata) in eastern North America. Am Midl Nat. 1998;140(2):325–42. doi: 10.1674/0003-0031(1998)140[0325:MSMODO]2.0.CO;2

70. Menz MHM, Brown BV, Wotton KR. Quantification of migrant hoverfly movements (Diptera: Syrphidae) on the west coast of North America. R Soc Open Sci. 2019;6: 190153. doi: 10.1098/rsos.190153

71. Buffalo V. Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin's Paradox. eLife. 2021;10:e67509. doi: 10.7554/eLife.67509

72. Altizer S, Hobson KA, Davis AK, De Roode JC, Wassenaar LI. Do healthy monarchs

migrate farther? Tracking natal origins of parasitized vs. uninfected monarch butterflies overwintering in Mexico. Breuker CJ, editor. PLOS ONE. 2015;10(11):e0141371. doi: 10.1371/journal.pone.0141371

73. Du B, Ding D, Ma C, Guo W, Kang L. Locust density shapes energy metabolism and oxidative stress resulting in divergence of flight traits. Proc Natl Acad Sci. 2022;119(1):e2115753118. doi: 10.1073/pnas.2115753118
74. Gatehouse AG. Insect migration: Variability and success in a capricious environment. Popul Ecol. 1994;36(2):165–71. doi: 10.1007/BF02514932
75. Chapman BB, Hulthén K, Blomqvist DR, Hansson LA, Nilsson JÅ, Brodersen J, et al. To boldly go: Individual differences in boldness influence migratory tendency. Ecol Lett. 2011;14(9):871–6. doi: 10.1111/j.1461-0248.2011.01648.x
76. Zhan S, Zhang W, Niitepõld K, Hsu J, Haeger JF, Zalucki MP, et al. The genetics of monarch butterfly migration and warning colouration. Nature. 2014;514:317–21. doi: 10.1038/nature13812
77. Van Petegem KHP, Pétillon J, Renault D, Wybouw N, Van Leeuwen T, Stoks R, et al. Empirically simulated spatial sorting points at fast epigenetic changes in dispersal behaviour. Evol Ecol. 2015;29:299–310. doi: 10.1007/s10682-015-9756-9
78. Boman J, Zhu Y, Höök L, Vila R, Talavera G, Backström N. Environmental stress during larval development induces DNA methylation shifts in the migratory painted lady butterfly (*Vanessa cardui*). Mol Ecol. 2023;00:1-11. doi: 10.1111/mec.16957
79. Näsvall K, Shipilina D, Vila R, Talavera G, Backstrom N. Resource availability affects activity profiles of regulatory elements in a long-distance butterfly migrant [preprint]. 2023. doi: 10.22541/au.167827909.99815237/v1
80. Jones CM, Papanicolaou A, Mironidis GK, Vontas J, Yang Y, Lim KS, et al. Genomewide transcriptional signatures of migratory flight activity in a globally invasive insect pest. Mol Ecol. 2015;24(19):4901–11. doi: 10.1111/mec.13362
81. Stefanescu C, Ubach A, Wiklund C. Timing of mating, reproductive status and resource availability in relation to migration in the painted lady butterfly. Animal Behaviour.

2021;172:145–53. doi: 10.1016/j.anbehav.2020.12.013

82. Hobson KA, Plint T, Serrano EG, Alvarez XM, Ramirez I, Longstaffe FJ. Within-wing isotopic (δ^2H, δ^{13}C, δ^{15}N) variation of monarch butterflies: Implications for studies of migratory origins and diet. Anim Migr. 2017;4:8–14. doi: 10.1515/ami-2017-0002

83. Paritte JM, Kelly JF. Effect of cleaning regime on stable-isotope ratios of feathers in Japanese Quail (*Coturnix japonica*). The Auk. 2009;126(1):165–74. doi: 10.1525/auk.2009.07187

84. Wassenaar LI, Hobson KA. Comparative equilibration and online technique for determination of non-exchangeable hydrogen of keratins for use in animal migration studies. Isotopes Environ Health Stud. 2003;39(3):211–7. doi: 10.1080/1025601031000096781

85. Soto DX, Koehler G, Wassenaar LI, Hobson KA. Re-evaluation of the hydrogen stable isotopic composition of keratin calibration standards for wildlife and forensic science applications. Rapid Commun Mass Spectrom. 2017;31(14):1193–203. doi: 10.1002/rcm.7893

86. Coplen TB, Qi H. USGS42 and USGS43: Human-hair stable hydrogen and oxygen isotopic reference materials and analytical methods for forensic science and implications for published measurement results. Forensic Sci Int. 2012;214:135–41. doi: 10.1016/j.forsciint.2011.07.035

87. Moore LJ, Murphy TJ, Barnes IL, Paulsen P. J. Absolute isotopic abundance ratios and atomic weight of a reference sample of strontium. J Res Natl Bur Stand. 1982;87(1):1–8.

88. Ma C, Vander Zanden HB, Wunder MB, Bowen GJ. assignR: An R package for isotope-based geographic assignment. Methods Ecol Evol. 2020;00:1–6. doi: 10.1111/2041-210x.13426

89. R Core Team. R: A language and environment for statistical computing [software]. Vienna, Austria: R Foundation for Statistical Computing. 2021. Available from: https://www.R-project.org/

90. Ghouri S, Reich MS, Lopez-Mañas R, Talavera G, Bowen G, Vila R, et al. A hydrogen isoscape for tracing the migration of terrestrial herbivorous insects across the Afro-Palearctic range [Preprint]. Authorea. 2023. doi: 10.22541/au.168897547.72773944/v1

91. Reich, MS, Ghouri S, Zabudsky S, Hu L, Le Corre M, Ng'iru I, Benyamini D, Shipilina S, Collins SC, Martins DJ, Vila R, Talavera G, Bataille CP. Trans-Saharan migratory patterns in Vanessa cardui and evidence for a southward leapfrog migration. [In preparation].

92. Vander Zanden HB, Nelson DM, Wunder MB, Conkling TJ, Katzner T. Application of isoscapes to determine geographic origin of terrestrial wildlife for conservation and management. Biol Conserv. 2018;228:268–80. doi: 10.1016/j.biocon.2018.10.019

93. Hallworth MT, Marra PP, McFarland KP, Zahendra S, Studds CE. Tracking dragons: Stable isotopes reveal the annual cycle of a long-distance migratory insect. Biol Lett. 2018;14(12). doi: 10.1098/rsbl.2018.0741

94. Katzner TE, Nelson DM, Braham MA, Doyle JM, Fernandez NB, Duerr AE, et al. Golden Eagle fatalities and the continental-scale consequences of local wind-energy generation. Conserv Biol. 2017;31(2):406–15. doi: 10.1111/cobi.12836

95. Rohlf FJ. tpsDig2 ver. 2.32. [software]. Ecology & Evolution, SUNY, Stony Brook; 2016.

96. Escobar-Suárez S, Villalobos-Leiva A, Fabres A, Órdenes-Clavería R, Cruz-Jofré F, Laroze D, et al. A geometric morphometrics and genetics characterization of *Vanessa carye* in an extreme elevational gradient in the Chilean Altiplano. Zool Anz. 2023;304:105–12. doi: 10.1016/j.jcz.2023.04.001

97. Soule AJ, Decker LE, Hunter MD. Effects of diet and temperature on monarch butterfly wing morphology and flight ability. J Insect Conserv. 2020;24(6):961–75. doi: 10.1007/s10841-020-00267-7

98. Adams DC, Otárola-Castillo E. geomorph: An R package for the collection and analysis of geometric morphometric shape data. Methods Ecol Evol. 2013;4:393–9. doi: 10.1111/2041-210X.12035

99. Garcia M, Juhos S, Larsson M, Olason PI, Martin M, Eisfeldt J, et al. Sarek: A portable

workflow for whole-genome sequencing analysis of germline and somatic variants. F1000Research. 2020;9(63):1-20. doi: 10.12688/f1000research.16665.2

100. Andrews S. FastQC: A quality control tool for high throughput sequence data [software]. 2010.
101. Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta. 2023;2:e107. doi: 10.1002/imt2.107
102. Chen S, Zhou Y, Chen Y, Gu J. fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90. doi: 10.1093/bioinformatics/bty560
103. Lohse K, Wright C, Talavera G, García-Berro A. The genome sequence of the painted lady, *Vanessa cardui* Linnaeus 1758. Wellcome Open Res. 2021;6:324. doi: 10.12688/wellcomeopenres.17358.1
104. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [preprint]. arXiv. 2013. Available from: http://arxiv.org/abs/1303.3997
105. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9. doi: 10.1093/bioinformatics/btp352
106. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491–8. doi: 10.1038/ng.806
107. Van Der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Curr Protoc Bioinforma. 2013;11(1110):1-43. doi: 10.1002/0471250953.bi1110s43
108. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27(21):2987–93. doi: 10.1093/bioinformatics/btr509
109. Shipilina D, Nasvall K, Höök L, Vila R, Talavera G, Backström N. Linkage mapping and genome annotation give novel insights into gene family expansions and regional

recombination rate variation in the painted lady (*Vanessa cardui*) butterfly. Genomics. 2022;114(6):110481. doi: 10.1101/2022.04.14.488360

110. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8. doi: 10.1093/bioinformatics/btr330

111. Privé F, Luu K, Vilhjálmsson BJ, Blum MGB. Performing highly efficient genome scans for local adaptation with R package pcadapt Version 4. Rosenberg M, editor. Mol Biol Evol. 2020;37(7):2153–4. doi: 10.1093/molbev/msaa053

112. Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. Fast and efficient estimation of individual ancestry coefficients. Genetics. 2014;196(4):973–83.

113. Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: Variational inference of population structure in large SNP data sets. Genetics. 2014;197(2):573–89. doi: 10.1534/genetics.113.160572

114. Korunes KL, Samuk K. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. Mol Ecol Resour. 2021;21(4):1359–68. doi: 10.1111/1755-0998.13326

115. Samuk K, Korunes K, Trevisani MD, Moshiri N. ksamuk/pixy: pixy 1.2.7.beta1 [software]. 2022. Available from: https://zenodo.org/record/6551490