

Migratory behavior is positively associated with genetic diversity in butterflies

Aurora García-Berro¹, Venkat Talla², Roger Vila³, Hong Kar Wai^{4,5}, Daria Shipilina^{2,6}, Kok Gan Chan^{5,7,8}, Naomi E. Pierce⁹, Niclas Backström², Gerard Talavera^{1,9,*}

¹Institut Botànic de Barcelona (IBB), CSIC-Ajuntament de Barcelona, Passeig del Migdia s/n, 08038 Barcelona, Catalonia, Spain, ²Department of Ecology and Genetics, Program of Evolutionary Biology, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden, ³ Institut de Biologia Evolutiva (CSIC-UPF), Passeig Marítim de la Barceloneta 37, 08003 Barcelona, Spain, ⁴Novel Bacteria and Drug Discovery Research Group (NBDD) and Microbiome and Bioresource Research Strength (MBRS), Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia, ⁵Division of Genetics and Molecular Biology, Institute of Biological Sciences, University of Malaya, Kuala Lumpur, Malaysia, ⁶Swedish Collegium for Advanced Study, Thunbergsvägen 2, 752 36, Uppsala, Sweden, ⁷International Genome Centre, Jiangsu University, Zhenjiang, China, ⁸Guangdong Provincial Key Laboratory of Marine Biology, Institute of Marine Sciences, Shantou University, Shantou, 515063, China, ⁹Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, USA

Abstract

Migration is typically associated with risk and uncertainty at the population level, but little is known about its cost and benefit trade-offs at the species level. Migratory insects often exhibit strong demographic fluctuations due to local bottlenecks and outbreaks. Here, we use genomic data to investigate levels of heterozygosity and long-term population size dynamics in migratory insects, as an alternative to classical local and short-term approaches such as regional field monitoring. We analyze whole-genome sequences from 97 Lepidoptera species and show that migratory species have significantly higher levels of genome-wide heterozygosity, a proxy for effective population size, than non-migratory species. Also, we contribute whole-genome data for one of the most emblematic insect migratory species, the painted lady butterfly (*Vanessa cardui*), sampled across its worldwide distribution range. This species exhibits one of the highest levels of heterozygosity described in Lepidoptera (2.95% ± 0.15). Coalescent modeling (PSMC) shows historical demographic stability in *V. cardui*, and high effective population size estimates of 2 to 20 million individuals 10,000 years ago. The study reveals that the high risks associated with migration and local environmental fluctuations do not seem to decrease overall genetic diversity and demographic stability in migratory Lepidoptera. We propose a “compensatory” demographic model for migratory r-strategist organisms in which local bottlenecks

are counterbalanced by reproductive success elsewhere within their typically large distributional ranges. Our findings highlight that the delimitations of populations are substantially different for sedentary and migratory insects, and that, in the latter, local and even regional field monitoring results may not reflect whole population dynamics. Genomic diversity patterns may elucidate key aspects of an insect’s migratory nature and population dynamics at large spatiotemporal scales.

Keywords: Genetic diversity, *Vanessa cardui*, Effective population size, Demography, Insect migration, PSMC, Population bottlenecks, Outbreaks, Migratory connectivity

Introduction

Animal movement plays an essential role in configuring species’ population dynamics. Among different types of movement observed in nature, migratory behavior results in seasonal long-distance movements at the whole population scale (Dingle, 1996; Dingle & Drake, 2007; Jonzén et al, 2011). Quantifying the costs and benefits of migration is a challenge and, consequently, the trade-offs of this strategy are not well understood. A key question that remains is: What are the long-term net effects of migration on a species’ population size and genetic diversity? On the one hand, migration can be beneficial since it allows for rapid exploitation of alternative habitats in response to environmental changes (Chapman et al., 2015). On the other hand, migration is also associated with substantial risks that can lead to large fluctuations in offspring number between successive generations. Costs are typ-

ically caused by increased predation, exhaustion, disorientation and unpredictable weather during the migration, or unsuitable habitat at the destination, all of which can lead to high mortality compared to sedentary species, and thus severe decreases in population size or extinctions for particular migratory clusters (Newton, 2008; Wilcove & Wikelski, 2008; Calvert et al., 2009; Chapman et al., 2015; Møller, 2008; Saunders et al., 2019).

Migratory animals can also experience long-term population bottlenecks at the species' range level. Since migratory ranges usually cover wide geographic areas, species can suffer from both long-term effects of climate change and habitat transformations in any of these areas, such as food shortage at critical breeding or stopover sites and/or increased exposure to diseases (Cox, 2010). It is therefore crucial to distinguish between bottlenecks/fluctuations that occur at a local and/or regional level, which are normally short-term, affecting individual migratory clusters within larger populations, and bottlenecks that occur at the whole population or species' range level, which are usually spread over many generations and caused by dramatic events at a large spatial scale (England et al., 2003). Population declines have been observed in a range of migratory birds and other vertebrates (Sanderson et al., 2006; Berger et al., 2008; Moller et al., 2008; Both et al., 2010; Wilson et al., 2018), some of which have resulted in extinction (Hung et al., 2014; Bairlein, 2016; Murray et al., 2017), indicating that migratory species may be particularly sensitive to ongoing climatic and habitat changes and thus of particular concern for conservation (Wilcove & Wikelski, 2008).

The costs of long-range migration in insects have been little studied (Chapman et al., 2015). There is no fixed set of expectations of how movement behavior may shape species' genetic diversity, as these may differ due to a variety of movements occurring in nature (Dingle, 1996, 2001; Dingle & Drake, 2007; Jonzén et al., 2011). The overall effective population size and the intensity of demographic fluctuations play large roles in the build-up and maintenance of genetic variation (Motro & Thomson, 1982; Whitlock, 1992). Migratory insects often exhibit vast population numbers (Chapman et al., 2011; Hu et al., 2016) and cyclic outbreaks (Wilson & Gatehouse, 1993; Chapuis et al., 2009; Satterfield et al., 2020). They also present some distinctive particularities in comparison with vertebrates that may shape demographic dynamics (Gao et al., 2020). These can include a higher reproductive potential, shorter generation times, or annual migrations that are completed through multiple generations. Migratory behavior may also counteract some of the factors that typically control insect population numbers, and result in comparatively larger population

sizes in migratory compared to non-migratory species. First, the migratory strategy allows escape from deteriorating environmental conditions. As many migratory insects breed continuously without diapausing, different generations can exploit a succession of favorable breeding grounds that may reduce the probability for local bottlenecks. The high mobility associated with long-range migration may also lead to unusually large distribution ranges, which can increase the chances of finding optimal conditions. Finally, multigenerational migration allows insects to move into enemy-free space, most importantly escaping from parasitoids and predators that often increase in numbers in breeding areas as the season advances (Altizer et al., 2011; Bartel et al., 2011; Stefanescu et al., 2012; Chapman et al., 2015).

In contrast, theory states that gene flow will have a positive effect on the global level of genetic diversity, the effect being dependent on the migration rate and the level of diversity among immigrants (Leblois et al., 2004, 2006; Chapuis et al., 2009, 2014; Pfeiler & Markow, 2017). In migratory Lepidoptera, it is thus expected that seasonal migrations increase genetic diversity within migratory ranges due to throughout mixing between migratory groups, although ultimately depending on the degree of migratory connectivity between them (Webster et al., 2002; Gao et al., 2020). However, genetic diversity may also vary in non-migratory Lepidoptera due to other movement behaviours, which we summarise in Table 1. Since migratory and non-migratory species likely evolved under different selective pressures, we might expect that they will show distinctive patterns in the regularity and scale of their movements, and therefore in their population dynamics (Dingle, 2001; Webster et al., 2002; Dingle & Drake, 2007; Bhaumik & Kunte, 2020) (Table 1).

Overall, we hypothesize that 1) a migratory life-style in Lepidoptera allows for a larger population size and, consequently, higher levels of genetic diversity in comparison to non-migratory species, and, 2) that demographic fluctuations between successive generations or seasons do not lead to severe bottlenecks or reduced genetic diversity at the species level. Here, we investigate whether migratory species of Lepidoptera show higher heterozygosity – a proxy for genetic diversity and effective population size – than non-migratory species. We estimate genomic heterozygosity across 381 individuals in 184 species in Lepidoptera, an order of insects whose life histories are reasonably well-known, and include a remarkable percentage of migratory species (ca. 3% found in butterflies; Chowdhury et al., 2021). As a paradigmatic case, we study in detail the demographic dynamics of the painted lady butterfly, *Vanessa cardui*. We generate whole-genome data and use coalescent

modeling to infer population size and demographic dynamics across worldwide populations. This species is the most cosmopolitan of all butterflies (Shields, 1992), and performs annual latitudinal round-trip migrations involving 8-10 generations without diapausing (Talavera et al., 2018; Menchetti et al., 2019). Various studies have documented dramatic local fluctuations in *V. cardui*, both within and between years, seemingly coupled to natural climate oscillations (Williams, 1970; Pollard et al., 1998; Asher et al., 2001; Vandenbosch, 2003; Stefanescu et al., 2013; Menchetti et al., 2019; Hu et al., 2021; López-Mañas et al., 2022). We investigate whether these local fluctuations shape the demographic history of the species at a global scale, impacting the overall species genetic diversity, or alternatively, whether the migratory strategy of *V. cardui* reduces the risk of severe bottlenecks and thus leads to maintenance of high levels of genetic diversity.

Materials and Methods

Heterozygosity estimates

Heterozygosity, or heterozygosity rate, is the fraction of nucleotides within an individual that differ between the chromosomes they inherit from their parents, a basic genomic feature that can be computationally complex to estimate when genome assemblies are not available and sequence data is limited to unassembled reads (Bryc, 2013). We estimated levels of heterozygosity for 381 individuals belonging to 184 Lepidoptera species (Table S1, Figure S2) using GenomeScope v1.0 (Vurture et al., 2017), a method based on the statistical analysis of the *k-mer* profile (occurrence of substrings of length *k*) of raw short read sequencing data (Vurture et al., 2017). Homozygous genomes have a simple Poisson profile while heterozygous genomes have a characteristic bimodal profile (Kajitani et al., 2014). GenomeScope attempts in several rounds to measure genome-wide heterozygosity by fitting a mixed model that accounts for the distortion that sequencing errors, repeats and read duplications exert in these profiles (Vurture et al., 2017).

All species (including multiple individuals when available) with >1 Gb short reads available in the SRA archive in the NCBI database (accessed June 2021) were included in the analysis. A few individuals from laboratory lines were discarded to avoid biased heterozygosity estimates as a consequence of potential inbreeding. The selected species included taxa from five butterfly and eight moth families, and represented a diversity of movement behaviors (Table S1). Reads were right-trimmed to be of the same length for each sample and filtered for a minimum quality of 20 using BB-Duk v38.61b (Bushnell, BBTools, 2020). We obtained coverage histograms with Jellyfish v2.2.6 (Marçais &

Kingsford, 2011) using several *k-mer* lengths for each species (15, 17, 19, 21, 23, 25) to account for the variability in size and repetitiveness of the genomes. We provided these as input to GenomeScope to obtain genome-wide heterozygosity estimates and selected the best model for each species based on a minimum fit of 90%.

It is important to note that *k-mer* based heterozygosity estimates from a single sample are not comparable to absolute values of heterozygosity estimates obtained from SNP-based approaches involving multiple samples. As *k-mer* based heterozygosity is estimated using the number of mismatches between *k-mers*, this will be much closer to the estimates based on the number of segregating sites in the population, such as the Watterson estimator of genetic diversity (θ). Generally, estimates of heterozygosity based on GenomeScope might be slightly higher compared to those based on population estimates (Vurture et al., 2017).

Behavioral scoring

We compiled evidence on migratory behavior from the literature for all taxa included in the dataset (Table S2). We systematically searched in Google Scholar and in a variety of regional field guidebooks for descriptors typically used in literature for Lepidoptera to describe movements of certain impact at geographical scale: the keywords “migration”, “migratory”, “migrating”, “dispersal”, “long-distance” and “long-range”, associated with each species name (including taxonomic synonyms). We also used as a working framework a list of nearly 600 butterfly species found to show evidence of migratory movements based on an exhaustive literature review (Chowdhury et al., 2021). All species of butterflies with migratory behavior retrieved from our literature search were also listed in Chowdhury et al (2021).

We scored the species in the dataset as belonging to one of three categories: sedentary, migratory and dispersive movements. Migratory behavior may be difficult to distinguish from other types of movement (Dingle, 1996; Dingle & Drake, 2007; Jonzén et al., 2011), and migration in Lepidoptera has generally been diagnosed based on population-level outcomes. For example, 92% of the butterfly species listed as migratory in the review by Chowdhury *et al.* (2021) were based only on field observations of directional mass movement, but only 8% of the species matched additional criteria at the experimental level. This 8% of species also had higher numbers of literature matches related to field-based observations of mass movements. As a general rule-of-thumb, we interpreted the absence of literature matches as absence of evidence of being a highly mobile species, and therefore scored these

species as sedentary. Although sedentary (= resident) species may display round-trip movements within their home range in the search of resources (e.g. when foraging and/or exhibiting territorial behavior), these have a minor impact on the population level movement. In searching for literature matches, the information was carefully scrutinized for population-level evidence of migratory behavior and other types of movement (Dingle, 1996). Our rationale is summarized in Table 1.

Briefly, we define migratory species as those exhibiting stimuli-driven, endogenously programmed, and non-resource-oriented obligate movements outside their home range, involving seasonal shifts of whole populations. We define dispersal as the outcome of other types of movements that show a resource-oriented behavior or a facultative response to prevailing conditions involving exploration. Dispersal is generally unpredictable, leading to relevant geographical shifts, but ones where not all populations move. Our definition of dispersal may overlap, partially or completely, with the concept of ranging (sensu Dingle & Drake, 2007) or nomadism (sensu Jonzén et al., 2011). Generally, the information more commonly extracted from literature to distinguish migration from dispersal was related to the geographical scale of the movements (longer *vs* shorter ranges), the demographic scale (whole populations *vs* parts of the populations), the regularity (seasonal *vs* irregular recurrence) and the nature of the movement (obligate *vs* facultative). In most cases, species considered migratory were strongly supported by multiple publications and detailed ecology, while species considered dispersive showed evidence of important dispersal outcomes, but did not necessarily have consistent support from multiple sources, suggesting to us a better fit with species scored as dispersive rather than sedentary or migratory.

The only case where intraspecific behavioral variability was documented was in *Danaus plexippus*, where specimens from known migratory versus sedentary populations (Zhan et al., 2014; Freedman et al., 2020) were treated separately in downstream analyses. Overall, our behavioral categorization resulted in 16 species scored as migratory (17.64%), 69 as sedentary (65.66%), and 13 as dispersive (12.74%) (Figure 1, Table S1).

Comparative analyses

We built a phylogenetic framework including all species for which we previously inferred levels of genomic heterozygosity. We retrieved five of the most commonly used genes in Lepidoptera phylogenetics from GenBank (Table S3) and used BEAST2 (Bouckaert et al., 2014) to infer phylogenetic relationships.

The substitution model was set to GTR+I+G with six gamma rate categories for each marker and a coalescent tree prior was set. We ran two independent MCMC chains of 100 million generations each, sampling values every 10,000 steps, and convergence was inspected in Tracer v.1.6.

We analysed the relationship between movement behaviour and heterozygosity levels using a phylogenetic generalized least squares (pGLS) regression to account for phylogenetic relatedness, as implemented in the R package ‘nlme’ (Pinheiro et al. 2016). For species where heterozygosity was estimated for multiple specimens, the mean values in the pGLS analysis were used. We also estimated the strength of the phylogenetic signal by inferring Pagel’s λ (Pagel, 1999) and Blomberg’s K (Blomberg, Garland & Ives 2003) in the model variables using the *phylosig* function in the R package ‘phytools’ (Revell, 2012) (Table S4).

Other life history traits could shape species’ heterozygosity levels and therefore act as confounding factors. As genetic diversity has been found to correlate with body size in butterflies (Mackintosh et al., 2019; Dapporto et al., 2019) and body size may correlate with population size (White et al., 2007), we included wingspan estimates obtained from the literature (Table S5) as a covariate in the pGLS test. In addition, because hybridization can elevate heterozygosity estimates, we excluded four potentially hybridizing taxa from our dataset (Table S1; Dupuis & Sperling, 2015, 2016; Herrell & Hazel, 1995; Mavárez et al., 2006).

Vanessa cardui sampling and sequencing

We collected six *V. cardui* specimens from natural populations in five geographic regions: one specimen in Ethiopia (14U392; N 9° 23’ 54.096”, E 38° 49’ 25.323”; November 28, 2014), Namibia (15D327; N 17° 52’ 16.086”, E 19° 24’ 22.611”; April 21, 2015), Nepal (07F575; N 28° 25’ 0.635”, E 83° 49’ 2.114”; April 15, 2008), California, USA (15A205; N 38° 34’ 30.414”, O 121° 34’ 38.117”; March 24, 2015), and two specimens in Spain (15A708 and 15A710; N 40° 15’ 29.16”, O 1° 36’ 20.879”; May 22, 2015). Genomic DNA was extracted from the thorax using the Omniprep DNA extraction kit (G-Biosciences). We prepared two libraries for each specimen: a short-insert paired-end library (550 bp) from 2 g of DNA following TruSeq DNA PCR-Free LT sample preparation guide, and a mate pair library from 1 g of DNA following Nextera Mate Pair gel-free protocol. All libraries were sequenced at the genomic facilities of the University of Malaya (Malaysia) at 126 bp from both ends using Illumina HiSeq 2500 technology. The sequencing reads for all specimens have been deposited in the ENA (European Nucleotide Archive) database under accession numbers PRJEB43266.

Table 1 Expected population-level outcomes from different types of movement in Lepidoptera potentially shaping connectivity and, ultimately, species' genetic diversity.

	Sedentary	Dispersive (other movements: e.g. ranging, long-range foraging, nomadism)	Migratory
Nature of movement	Resource-oriented, not exploratory	Facultative response to prevailing conditions, exploratory, resource- oriented, terminated by suitable resources	Obligate, anticipatory, endogenously programmed (pre-emption), temporary inhibition of resource- driven response
Regularity	Irregular	Irregularly recurrent (between-year variability)	Seasonally recurrent (constant)
Proportion of the population involved	No displacement	Individual-scale: only part of the population may move	Population-scale: whole population moves
Distance between individuals after movements	Constant	Potentially increased	Nearly constant
Extent of movements (geographical scale)	Short, within habitat	Short (between habitats) to occasionally long- range (between regions), usually one-way	Long-range, between regions, usually two-way
Duration of movements	Very short	Short to long, but often interrupted	Long and sustained
Annual number of generations involved	All	None, one or more than one	All or nearly all
Seasonal home-range occupancy	Always occupied (not abandoned)	Always occupied (not abandoned)	Temporarily abandoned/colonized
Safety of movement	Predictable (but not secured)	Unpredictable safe habitat at destination	Predictable (but not secured) safe habitat at destination
Predictability/regularity of movement	Yes	No	Yes
Philopatry and migratory connectivity	High (site-specific)	Very low	Low (region-specific) to high (site-specific)

Vanessa cardui reference-based assemblies

Five re-sequenced samples, representatives from different geographic regions, were used for demographic inference. We obtained an average of 205 million reads from these samples (Table 2) and used BWA v0.7.12 (Li & Durbin, 2010) to map the reads to the available *V. cardui* genome (ilVanCard2.1, NCBI accession number GCA_905220365.1) (Lohse et al., 2022). An average of 88.6% of the reads were mapped and both mapping quality and mean depth of coverage across samples was high (Table 2), as assessed using Qualimap v2.2.1 (Okonechnikov, Conesa & García-Alcalde, 2016). Mapped sequences were sorted and indexed using SAMTOOLS v1.3 (Danecek et al., 2021). We also assessed the degree of completeness by analysis of the arthropod BUSCO v3.0.2 gene set (Simão et al., 2015) using gVolante (Nishimura, Hara & Kuraku, 2017). Of the 1013 genes queried, an average of 999 com-

plete genes (98.63%) were retrieved from the assemblies, showing a high degree of completeness.

Vanessa cardui demographic inference

We used the Pairwise Sequentially Markovian Coalescent (PSMC) (Li & Durbin, 2011) to investigate the demographic history of *V. cardui* samples from diverse geographical origins. This method relies on coalescence to infer past recombination and mutation events along the diploid genome of a single individual. As an input for PSMC, we used each individual's consensus sequence generated with BCFtools (Li, 2011). Variant calling was performed using the SAMTOOLS command *mpileup* (Li & Durbin, 2009). Repeat-rich regions are prone to bias effective population size (N_e) estimates due to uncertainty in the quality of the variant calls, their population-level dynamics and deviations from the neutral expectations (Patil & Vijay,

Table 2 Quality statistics of the five *V. cardui* genomes

	Europe	N. America	S. Africa	E. Africa	Asia
Number of reads	171,431,637	176,271,556	171,520,074	226,399,734	276,861,301
Mapped reads	160,199,554	163,200,113	122,289,727	208,310,949	259,227,222
Percentage of mapped reads	93.45	92.58	71.30	92.01	93.63
Mean mapping quality	48.60	48.32	48.19	48.29	47.92
Mean coverage	42.36	43.05	32.34	55.09	68.42
GC content (%)	32.88	33.05	33.17	32.72	33.41
Heterozygosity (%)	3.13	2.96	2.99	2.78	2.77
BUSCO complete genes (%)	999 (98.62%)	997 (98.42%)	1001 (98.82%)	999 (98.62%)	1000 (98.72%)
BUSCO complete + partial genes (%)	1004 (99.11%)	1003 (99.01%)	1006 (99.31%)	1005 (99.21%)	1005 (99.21%)

2021). Thus, we excluded repeats as well as coding sequences in order to take into account only neutral sites. We used BEDtools (Quinlan & Hall, 2010) to exclude these regions from the consensus sequence using the available *V. cardui* genome and repeat/transposable element annotation as a reference (GCA_905220365.1: Lohse et al., 2022; Shipilina et al., 2022). Following Nadachowska-Brzyska et al., (2016), we excluded sites for which the mapping coverage was $< 1/3$ of the average read depth to minimize erroneously called homozygous sites, and sites where read depth was higher than twice the average in order to minimize the impact of potential erroneously collapsed regions in the assembly. We also filtered out positions with *Phred* score < 20 . The upper limit of the TMRCA was set to 5 and parameter space for time intervals was set to “4+30*2+4+6+10”, so that at least 10 recombination events occurred after the twentieth iteration (Li & Durbin, 2011). In order to get estimates of variance in effective population size (N_e), we performed 100 PSMC runs by randomly sampling with replacement 5 Mb sequence segments obtained from the consensus genome sequence. Average generation time was set to 0.11 years, corresponding to 9 generations per year (Talavera & Vila, 2017). The mutation rate was set to 2.9×10^{-9} mutations per year, corresponding to the estimate from the closest relative butterfly species, *Heliconius melpomene* (Keightley et al., 2015).

Results

Heterozygosity rates in Lepidoptera

From the heterozygosity estimates performed for 183 Lepidoptera species (Table S1), we excluded 53 species for which the GenomeScope model failed to converge, and 33 species for which the minimum model fit was below 90%. Heterozygosity in the 97 remaining species ranged between 0.21 % in *Heliconius hermathena* and 3.62 % in *Junonia oenone* (Figure 1, Table S1). The average heterozygosity for the six specimens of *V. cardui* was 2.95 % [2.77 – 3.13 %] (Table S1). No significant correlation was found between heterozygosity values and *k-mer* coverage (Pearson’s correlation, $t =$

0.62039, P -value = 0.537), suggesting that *k-mer* coverage was not a factor influencing our heterozygosity estimates. For 48 species, we could infer heterozygosity values in more than one individual (19 species with 2, 21 species with 3, 7 species with 4, and *V. cardui* with 6 individuals). Average intra-specific standard deviation in heterozygosity was 0.22, the minimum was 0.001, and the maximum was 0.86, indicating that estimates based on single individuals were generally robust. Intra-specific deviation did not correlate with species’ average heterozygosity (Pearson’s correlation, $t = 0.62039$, P -value = 0.537) (Figure S1), suggesting a random distribution of this variability. Additionally, outliers showing highest intra-specific differences belonged to the genus *Papilio* (a group known to hybridize frequently) (Herrell & Hazel, 1995; Dupuis & Sperling, 2015, 2016) or to be used as laboratory model systems (e.g. *Agrotis ipsilon*), both potentially affecting individual heterozygosity estimates. However, we found no evidence of either hybridization or inbreeding in individuals of these taxa in the original studies, so none were excluded from the analyses.

Migratory species had significantly higher heterozygosity values compared to sedentary species (mean \pm s.e.: Migratory: 2.48 ± 0.65 , Sedentary: 1.55 ± 0.79 ; pGLS_{Strategy}: F -stat = 5.5978, P -value = 0.0200; Tukey Contrasts: P -value = 0.00514), and species categorized as dispersive (mean \pm s.e.: 1.30 ± 0.835 ; Tukey Contrasts: P -value = 0.00893) (Figure 1). No significant differences in heterozygosity were found between the categories sedentary and dispersive (Tukey Contrasts: P -value = 0.08911). Wingspan had no significant effect on heterozygosity as a fixed effect (pGLS_{Wingspan}: F -stat = 2.18108, P -value = 0.1429), although the interaction term was significant: wingspan and heterozygosity had a significant negative relationship in migratory species (pGLS_{Wingspan*Strategy}: F -stat = 28.672282, P -value < 0.0001). Wingspan was independent from the explanatory variable ‘movement strategy’ (ANOVA, F -stat = 7.998896, P -value = 0.0057).

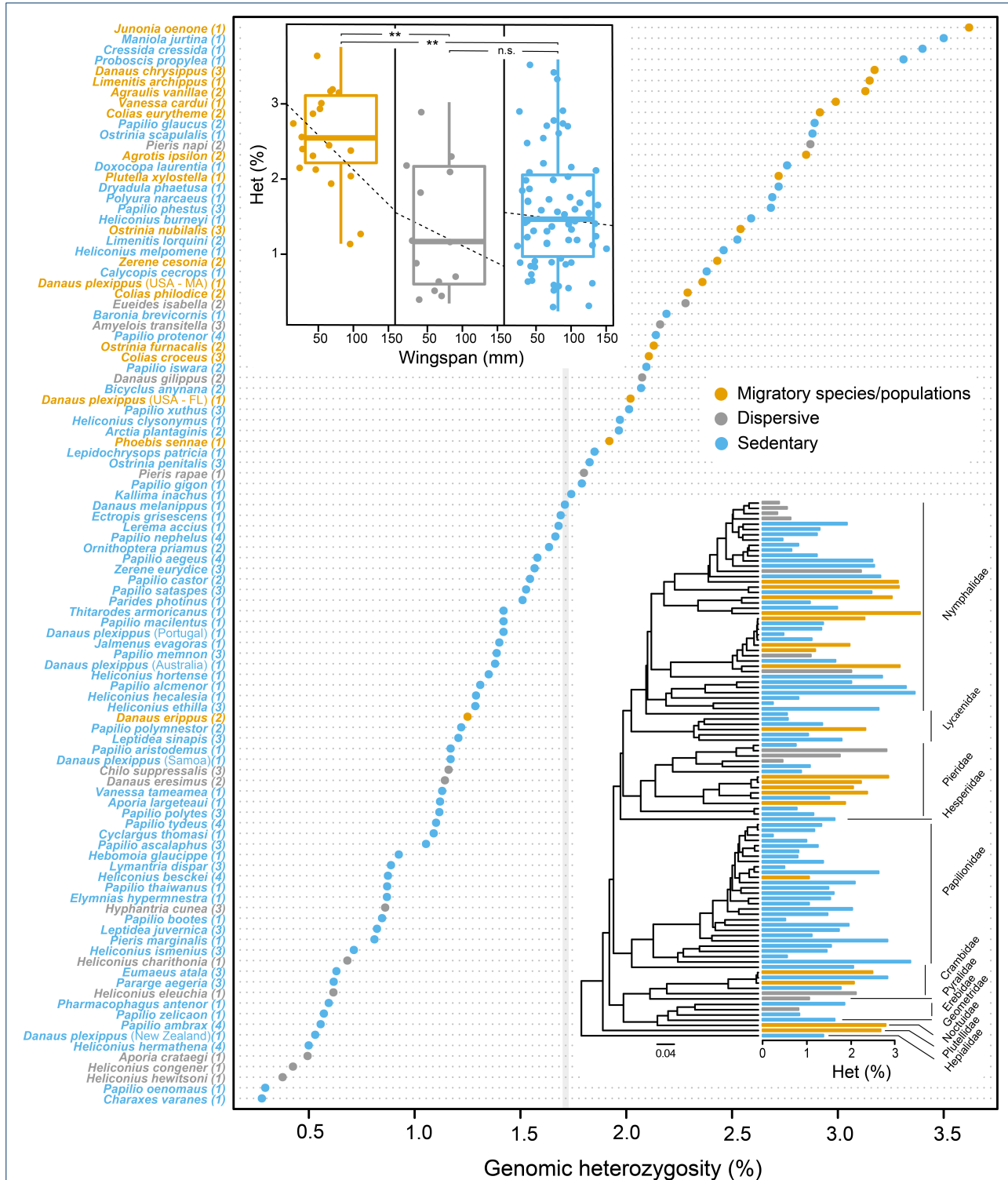


Figure 1 (a) Genomic heterozygosity for 185 individuals belonging to 97 Lepidoptera species calculated from short read Illumina reads using GenomeScope v1.0. Average estimates for species with several specimens are shown, with the exception of *Danaus plexippus*, for which migratory and non-migratory populations are illustrated. Numbers next to species names denote the number of individuals used for heterozygosity inference. The grey line indicates the average value for all analysed species. Species were classified as sedentary (in blue), migratory (in orange), and dispersive (in grey). (b) A pGLS analysis was used to test for significant deviation in heterozygosity between groups, accounting for wing size as proxy for body size (* = P -value < 0.05, ** = P -value < 0.01, *** = P -value < 0.001 and n.s. = non-significant). (c) Phylogenetic diversity of the sampled species. Bars next to tips show average heterozygosity estimates.

Heterozygosity estimates were phylogenetically independent as showed by low Blomberg's K and Pagel's λ , and migratory strategy showed weak phylogenetic signal, with λ and K values < 0.5 (Table S4). However, wingspan was distributed as expected under the Brownian motion (BM) model of trait evolution and exhibited strong phylogenetic dependence ($\lambda = 0.99$ and $K = 1.50$).

Vanessa cardui demographic history

The estimated change in effective population size (N_e) over time for five *V. cardui* individuals from the different sampled geographic regions is shown in Figure 2. The demographic histories of distant populations shared 1) a lack of abrupt N_e fluctuations and 2) a progressive population expansion that reaches a peak approximately ten thousand years ago (10 kya). We identify three phases in the demographic history of the different populations:

1) 150 – 60 kya: a coalescent time for all populations at approximately 150 kya. Ancestral effective population size estimates for the species ranged from 10,000 to 15,000 individuals.

2) 80 – 20 kya: a period of slow but steady increase in population size, reaching N_e values between 200,000 and 800,000 individuals. The trajectories and the bootstrap replicates remained overlapping among populations.

3) 20 – 10 kya: a period of stabilization and slight decrease in N_e , progressively differing among populations, suggesting an increase in continental-level population structure along with differences in N_e . Estimates of N_e for this period range from 10^6 to 10^7 effective individuals.

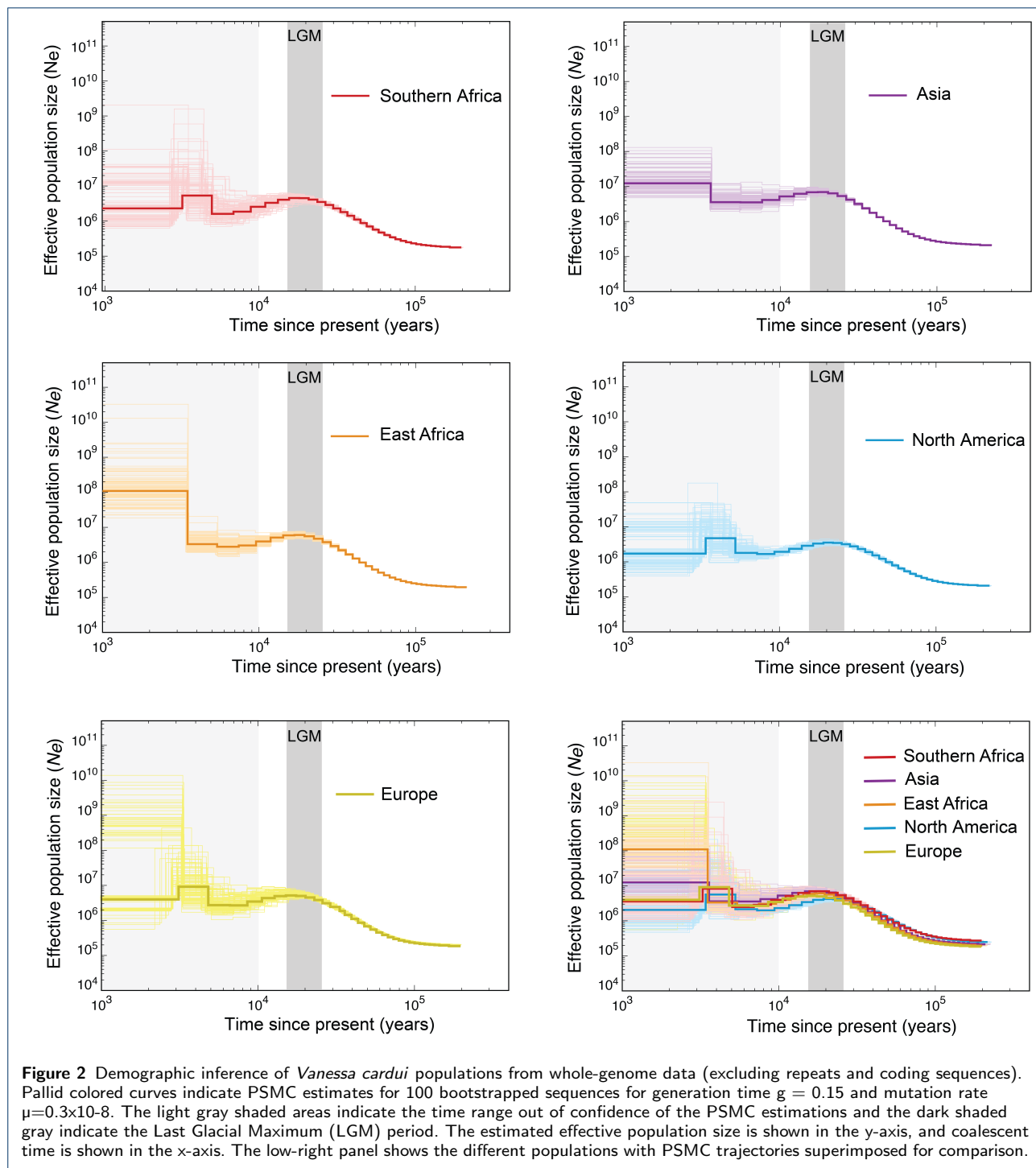
Discussion

Genomic heterozygosity, effective population size and migration

We show that migratory Lepidoptera have higher heterozygosity values than non-migratory species. The highest levels of heterozygosity are found in species that are extremely abundant and/or are habitat generalists, many of them ranked as migratory or dispersive. *Junonia oenone* (heterozygosity = 3.62 %), for example, is migratory and one of the most abundant and widespread butterflies in Africa, flying year-round (including during the dry season), and thus avoiding diapause. Similar cases include the butterfly species *Danaus chrysippus* (3.23 %), *Cotias eurytheme* (3.04 %), *Pieris napi* (2.76 %) and the moth species *Agrotis ipsilon* (3.34%), *Plutella xylostella* (2.72 %) and *Helicoverpa zea* (Anderson et al., 2018). Among the migratory species, *V. cardui* ($2.95\% \pm 0.15$) exhibits one of the highest levels of heterozygosity described in

Lepidoptera (Li et al., 2019), ranking among the top 10 % of the most heterozygous species in our dataset (Figure 1). Other species found among the most heterozygous (Table S1) are suggested to hybridize, a process that can potentially be the cause of the observed high levels of heterozygosity. This concerns for example *Papilio polyxenes*, *P. joanae*, *P. machaon* (Herrell & Hazel, 1995; Dupuis & Sperling, 2015, 2016) and *Heliconius heurippa* (Mavárez et al., 2006).

A positive association between genetic diversity at neutral sites and N_e is expected, since larger populations have a higher input of novel mutations and reduced genetic drift (Wright, 1931; Kimura, 1983; Nevo, 1978; Leffler et al., 2012; Mackintosh et al., 2019). Our results suggest that migratory butterflies tend to have comparatively large N_e . For example, in migratory monarch populations, N_e is inferred to be in the order of 10^7 (Zhan et al., 2014; Talla et al., 2020), compared to N_e of $10^5 - 10^6$ in sedentary populations (Zhan et al., 2014), the latter falling within a general range estimated for butterflies (Mackintosh et al., 2019). Our heterozygosity estimates for the monarch butterfly also followed a pattern of nearly twice the difference between behavioral categories: ca. 2 % in migratory and 1 ± 0.5 % in sedentary populations. For *V. cardui*, N_e were in the range of 10^6 to 10^7 at 10 kya. As census population size is expected to be larger than N_e (Fisher & Ford, 1947; Wright, 1948, Palstra & Ruzzante, 2008), one could debate whether such values are realistic. Several arguments support the existence of extraordinarily large population sizes in *V. cardui*. First, extremely large distribution ranges, typical of migratory insects, should facilitate elevated population sizes. The extreme dispersal capacity of *V. cardui* individuals is reflected in the vast migratory ranges that are covered annually across different continents. The species is also a polyphagous herbivore, and the diversity of larval hostplants and their phenology most likely expand opportunities for breeding success. As a typical r-strategist migratory insect, large population sizes may be advantageous in exploiting ephemeral flush vegetation, as well as in balancing the cost of higher mortality rates associated with migration (Dingle, 1978; Hill, 1983; Chapman et al., 2015). Moreover, the species has high reproductive potential (varying over several orders-of-magnitude in comparison to vertebrates), where single females may lay $> 1,000$ eggs (Benyamini, 2017), and can eventually behave as pests in particular breeding sites. Lastly, migration also facilitates higher population sizes by allowing escape from high parasitoid loads, an important factor exerting population control in insects (Jeffries & Lawton, 1984; Owen, 1987; Stefanescu et al., 2012; Chapman et al., 2015).



Although it is expected that highly abundant migratory species show high levels of heterozygosity given their greater opportunity to randomly share alleles between large populations, not all migratory species fall within this expectation. Indeed, there seem to be a few outliers that are not consistent with a correlation between genetic diversity and long-term population

size. In Mackintosh et al., 2019, estimates of neutral genetic diversity for some migratory butterflies were among the bottom 15 % of European species, such as *Vanessa atalanta*, *Pontia daplidice* or *Pieris brassicae*. In our dataset, an exception is *Danaus erippus* (1.16 %), which appears to have a heterozygosity level below average. A deviation from a strict linear relation-

ship between N_e and heterozygosity is not unexpected given a set of possible confounding factors. For example, recent geographic expansions and corresponding increases in population size, could lead to populations that have not yet reached mutation-drift equilibrium (Buffalo, 2021) – this has been suggested for example in *P. brassicae* and *P. rapae* (Mackintosh et al., 2019; Ryan et al., 2019). Another potential factor concerns the role that natural selection can play in decreasing neutral variation at physically linked sites (Ellegren & Galtier, 2016). In species with large population sizes, such ‘linked selection’ should be a more pervasive force since the efficiency of selection scales with N_e (Leffler et al., 2012; Corbett-Detig et al., 2015; Mackintosh et al., 2019; Talla et al., 2019). A paradigmatic case is the extinction of the passenger pigeon, once the most abundant migratory bird in North America and possibly in the world. Its surprisingly low genetic diversity has been suggested to be a collateral effect of natural selection (Murray et al., 2017). However, the effect of linked selection is also dependent on the recombination landscape, and detailed recombination maps will be needed to gain further insight into the relative effects of selection, drift and demography driving differences in heterozygosity between species with similar life histories. In *V. cardui*, PSMC results show no evidence that N_e drops despite the large population sizes that could enhance the action of selection, suggesting that selection has not been a major factor shaping the demography of this species.

Other potential factors shaping genomic heterozygosity involve migratory connectivity. Some species may have strong philopatry, resulting in high migratory connectivity between departure and destination, and thus little mixing between populations or migratory ranges (Battey et al., 2017; Gao et al., 2020), despite their long migratory movements. In the monarch butterfly migratory range, two (eastern and western) breeding and overwintering areas are repeatedly visited every year, and populations maintain their own phenotypic characteristics (Vanderbosch, 2003; Talla et al., 2020). This contrasts with the strategy of *V. cardui*, where individuals may be more opportunistic in their breeding destinations within each of the predictable suitable areas occupied in each multigenerational step (*i.e.* lower philopatry, in the sense that the populations do not rely on specific sites but rather on large regions). This results in lower migratory connectivity and thus higher allele sharing between different breeding sites.

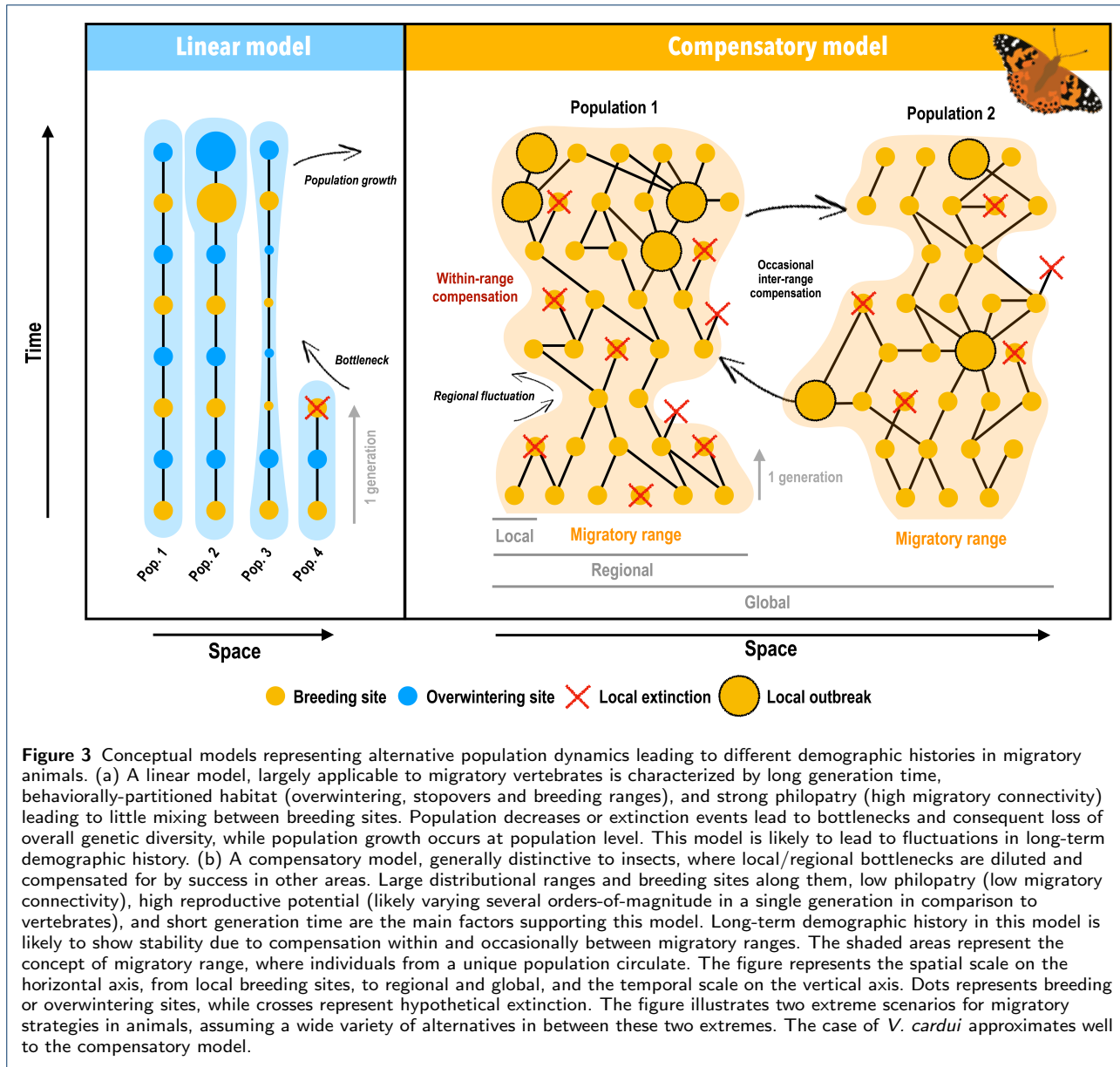
Lastly, diapause may also influence migratory species’ genetic diversity and population size, as the number of generations per year is higher in species that breed continuously year-round compared to diapausing species.

For example, the number of generations in *V. cardui* is twice that of *D. plexippus*, but also each *V. cardui* generation migrates and breeds in a new region, unlike the monarchs that keep three breeding generations within the same range (Flockhart et al., 2013). This uninterrupted movement and the high number of breeding regions/generations per year dramatically increase the possibilities of gene flow and is probably a relevant factor in explaining the very high levels of genomic heterozygosity in non-diapausing insects such as *V. cardui* and *J. oenone*.

The previously mentioned factors are expected to play a major role for migratory species, but these may not be observed in species that carry out other types of movements resulting in long-range dispersal. When dispersal is not seasonally recurrent and occurs at the individual level and not at the whole population scale, movement can increase the spatial distance between individuals, and eventually lower densities. While population admixture is expected to be large within the home range of these species, physical separation of individuals outside its range should limit the exchange of alleles and, eventually, may produce variation that is geographically structured. Species that can occasionally perform long-range dispersal have the potential to have larger distributional ranges than sedentary species, but not necessarily higher population size and heterozygosity, as our results suggest.

Short-term migratory cost but long-term demographic stability

Vanessa cardui demographic history shows a stable population growth over time, high effective population sizes, and no apparent dramatic fluctuations or bottlenecks overall. This is consistent across populations worldwide. Long-term stability contrasts with the observation of frequent short-term population fluctuations and regional outbreaks observed in migratory insects between successive generations and years (Vanderbosch, 2003; Chapuis et al., 2009; Chapman et al., 2011; Hu et al., 2016; López-Mañás et al., 2022). Chapuis et al., (2009) found that outbreaking periods in the migratory locust (*Locusta migratoria*) did not strongly affect genetic diversity over time, either in outbreaking or non-outbreaking populations. Theory predicts that long-term effective size in populations recurrently fluctuating in their size can be approximated by calculating the harmonic mean of population size over all point estimates during times of fluctuations (Vucetich et al., 1997; Wang, 2005). The reason for this is that the loss of genetic diversity during periods of comparatively small N_e is expected to be higher than the increase in genetic diversity (due to mutational input) during periods of comparatively



large N_e . This means that the long-term N_e should be closer to the remission periods rather than to the outbreaking periods (Motro & Thomson 1982; Chapuis et al., 2009). Under this scenario, an outbreak remission should not be considered a bottleneck because it affects census but not N_e . This rule probably applies to *V. cardui* and other migratory Lepidoptera, where regional fluctuations in census size may hardly affect the overall N_e of the species. Actually, outbreak events may lead to a homogenizing effect in the population structure of migratory insects (Chapuis et al., 2009; Anderson et al., 2018; Kaňuch et al., 2020). Given the large breeding ranges in most migratory insects, it is unlikely that local or even regional outbreaks or bottle-

necks affect most populations across the whole range synchronically. Thus, a dilution or compensation effect of local demographic fluctuations takes place at the species level because of the extremely wide range of populations (Figure 3). For example, in *V. cardui* populations apparently cover entire continents or even intercontinental systems, such as the Palearctic-Africa system, and outbreaks may occur occasionally anywhere within large monthly breeding ranges (López-Mañas et al., 2022). This scenario is probably different in philopatric migratory vertebrates though, where breeding or overwintering ranges are smaller, more delimited, and thus more vulnerable to suffer bottlenecks and population declines if their habitats are disturbed.

In this case, whole populations could be affected and genetic diversity loss is likely, with further population size recovery relying on the expansion of remaining populations (Figure 3).

Thus, a “compensatory” demographic model of migration in insects may result in a potential evolutionary advantage over the long-term, provided population sizes remain large enough compared to 1) sedentary species, where populations have limited ranges, thus making them potentially vulnerable to local habitat disturbance, unpredictable climatic events or parasitoid control, 2) strongly philopatric species, where success in different parts of the cycle rely on resources in specific sites, or 3) diapausing migratory species, where eventual bottlenecks could not be avoided during overwintering if important gatherings occur at a few key spots that may be exposed to environmental pressure.

Delimiting populations in migratory insect species

Migratory ranges (or circuits) in insects could be defined as geographical ranges, even if parts of the ranges are only used temporarily, where gene flow is higher than between ranges. Migration in *V. cardui* follows a recurrent latitudinal progression that delineate a set of predictable multigenerational steps (Stefanescu et al., 2013; Menchetti et al., 2019), although gene flow may occur longitudinally to some extent. Thus, migratory ranges in this species might be generally defined as large longitudinal blocks where latitudinal gene flow is higher than longitudinal gene flow. However, the global configuration of these migratory ranges in *V. cardui* is not yet well understood. The species shares similar worldwide patterns of N_e variation over time, despite slight differences in North American and Southern African populations. This probably explains why present-day geographical populations share a common demographic history, implying that *V. cardui* is nearly panmictic, as found in other highly mobile species (Lyons et al., 2012; Raymond et al., 2013; Cristofari et al., 2016; Pfeiler and Markow, 2017; Talla et al., 2020), or at least that populations may circulate within extremely large distributional ranges. Taken together, our findings highlight the need to revisit the use of the population concept in *V. cardui* and in migratory insects in general. Dedicated population genetic studies are needed to infer historical and current patterns of gene flow and properly delimit migratory ranges.

Acknowledgements

We thank Krystyna Nadachowska-Brzyska and Leonardo Dapporto for helpful methodological advice. This work was funded by the National Geographic Society (grant WW1-300R-18) and by the grant PID2020-117739GA-I00 from MCIN / AEI / 10.13039/501100011033 to G.T., by fellowship FPU19/01593 to A.G.-B., by grants from the Putnam

Expeditionary Fund of the Museum of Comparative Zoology to G.T. and N.E.P., by projects PID2019-107078GB-I00 / MCIN / AEI / 10.13039/501100011033 and 2017-SGR-991 (Generalitat de Catalunya) to R.V. and G.T., by the University of Malaya (grant H50001-A-000027) to K.G.C., by the Swedish Collegium for Advanced Science (SCAS; Natural Sciences Programme, Knut and Alice Wallenberg Foundation, Postdoc funding) to D.S., by NSF grant DEB-1541560 to N.E.P., and by the Swedish Research Council FORMAS (grant 2019-00670) to N.B., G.T., and R.V. The authors also 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.

Availability of data

Sequencing reads have been deposited in the ENA (European Nucleotide Archive) database under accession number PRJEB43266.

Authors' contributions

A.G.-B., R.V., N.B. and G.T. conceived the study. G.T. and R.V. collected samples. H.K.W. and G.T. carried out laboratory work. A.G.-B., V.T. and G.T. analysed data. A.G.-B. and G.T. wrote the first version of the manuscript. All authors contributed in interpreting results and edited and approved the final version of the manuscript.

ORCID

Aurora García-Berro <https://orcid.org/0000-0002-2419-2516>

Venkat Talla <http://orcid.org/0000-0003-2653-6770>

Roger Vila <http://orcid.org/0000-0002-2447-4388>

Hong Kar Wai <http://orcid.org/0000-0002-3489-9056>

Daria Shipilina <http://orcid.org/0000-0002-1145-9226>

Kok Gan Chan <http://orcid.org/0000-0002-1883-1115>

Naomi E. Pierce <http://orcid.org/0000-0003-3366-1625>

Niclas Backström <http://orcid.org/0000-0002-0961-8427>

Gerard Talavera <http://orcid.org/0000-0003-1112-1345>

References

- Altizer, S., Bartel, R., & Han, B. A. (2011). Animal migration and infectious disease risk. *Science*, 331, 296–302. <https://doi.org/10.1126/science.1194694>
- Anderson, C. J., Oakeshott, J. G., Tay, W. T., Gordon, K. H. J., Zwick, A., & Walsh, T. K. (2018). Hybridization and gene flow in the mega-pest lineage of moth, *Helicoverpa*. *Proceedings of the National Academy of Sciences of the United States of America*, 115(19), 5034–5039. <https://doi.org/10.1073/pnas.1718831115>
- Asher, J., Warren, M., Fox R., Harding P., Jeffcoate, G., & Jeffcoate, S. (2001). *The millennium atlas of butterflies in Britain and Ireland*. Oxford University Press.
- Bairlein, F. (2016). Migratory birds under threat. *Science*, 354(6312), 547–48. <https://doi.org/10.1126/science.aah6647>.
- Bartel, R. A., Oberhauser, K. S., De Roode, J. C., & Altizer, S. M. (2011). Monarch butterfly migration and parasite transmission in eastern North America. *Ecology*, 92(2), 342–351. <https://doi.org/10.1890/10-0489.1>
- Battey, C. J., Linck, E. B., Epperly, K. L., French, C., Slager, D. L., Sykes Jr, P. W., & Klicka, J. (2018). A migratory divide in the Painted Bunting (*Passerina ciris*). *The American Naturalist*, 191(2), 259–268. <https://doi.org/10.1086/695439>
- Benyamini, D. (2017). A swarm of millions of *Vanessa cardui* (Linnaeus, 1758) in winter-spring 2015-2016 in the south-east Mediterranean – The missing link. *Atalanta*, 48, 103–128.
- Bushnell, B. (2020). BBDuk. Jt Genome Inst. Available online: <https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-userguide/bbd-duk-guide/> (accessed on 25 June 2020).
- Berger, J., Young, J. K., & Berger, K. M. (2008). Protecting migration corridors: challenges and optimism for Mongolian saiga. *PLoS biology*, 6(7), e165. <https://doi.org/10.1371/journal.pbio.0060165>

- Bhaumik, V., & Kunte, K. (2020). Dispersal and migration have contrasting effects on butterfly flight morphology and reproduction. *Biology letters*, 16(8), 20200393. <https://doi.org/10.1098/rsbl.2020.0393>
- Blomberg, S. P., Garland, T., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, 57(4), 717–745. <https://doi.org/10.1111/j.0014-3820.2003.tb00285.x>
- Both, C., Van Turnhout, C. A., Bijlsma, R. G., Siepel, H., Van Strien, A. J., & Foppen, R. P. (2010). Avian population consequences of climate change are most severe for long-distance migrants in seasonal habitats. *Proceedings of the Royal Society B: Biological Sciences*, 277(1685), 1259–1266. <https://doi.org/10.1098/rspb.2009.1525>
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology*, 10(4), e1003537. <https://doi.org/10.1371/journal.pcbi.1003537>
- Bryc, K., Patterson, N., & Reich, D. (2013). A novel approach to estimating heterozygosity from low-coverage genome sequence. *Genetics*, 195(2), 553–561. <https://doi.org/10.1534/genetics.113.154500>
- Buffalo, V. (2021). Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin's Paradox. *Elife*, 10, e67509. <https://doi.org/10.7554/eLife.67509>
- Calvert, A. M., Walde, S. J., & Taylor P. D. (2009). Nonbreeding-season drivers of population dynamics in seasonal migrants: Conservation parallels across taxa. *Avian Conservation and Ecology* 4(2). <https://doi.org/10.5751/ace-00335-040205>
- Chapman, J. W., Drake, V. A., & Reynolds, D. R. (2011). Recent insights from radar studies of insect flight. *Annual Review of Entomology*, 56(1), 337–356. <https://doi.org/10.1146/annurev-ento-120709-144820>
- Chapman, J. W., Reynolds, D. R., & Wilson, K. (2015). Long-range seasonal migration in insects: mechanisms, evolutionary drivers and ecological consequences. *Ecology Letters* 18(3), 287–302. <https://doi.org/10.1111/ele.12407>
- Chapuis, M. P., Loiseau, A., Michalakakis, Y., Lecoq, M., Franc, A., & Estoup, A. (2009). Outbreaks, gene flow and effective population size in the migratory locust, *Locusta migratoria*: A regional-scale comparative survey. *Molecular Ecology*, 18(5), 792–800. <https://doi.org/10.1111/j.1365-294X.2008.04072.x>
- Chapuis, M. P., Plantamp, C., Blondin, L., Pagès, C., Vassal, J. M., & Lecoq, M. (2014). Demographic processes shaping genetic variation of the solitary phase of the desert locust. *Molecular Ecology*, 23(7), 1749–1763. <https://doi.org/10.1111/mec.12687>
- Chowdhury, S., Fuller, R. A., Dingle, H., Chapman, J. W., & Zalucki, M. P. (2021). Migration in butterflies: a global overview. *Biological Reviews*, 96(4), 1462–1483. <https://doi.org/10.1111/brv.12714>
- Corbett-Detig, R. B., Hartl, D. L., & Sackton, T. B. (2015). Natural selection constrains neutral diversity across a wide range of species. *PLoS biology*, 13(4), e1002112. <https://doi.org/10.1371/journal.pbio.1002112>
- Cox, G. W. (2010). *Bird migration and global change*. Island Press.
- Cristofari, R., Bertorelle, G., Ancel, A., Benazzo, A., Le Maho, Y., Ponganis, P. J., ... Trucchi, E. (2016). Full circumpolar migration ensures evolutionary unity in the Emperor penguin. *Nature Communications*, 7, 11842. <https://doi.org/10.1038/ncomms11842>
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan V., Pollard, M.O., ... & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008. <https://doi.org/10.1093/gigascience/giab008>
- Dapporto, L., Cini, A., Vodř, R., Dincă, V., Wiemers, M., Menchetti, M., ... & Vila, R. (2019). Integrating three comprehensive data sets shows that mitochondrial DNA variation is linked to species traits and paleogeographic events in European butterflies. *Molecular Ecology Resources*, 19(6), 1623–1636. <https://doi.org/10.1111/1755-0998.13059>
- Dingle, H. (1978). Migration and diapause in tropical, temperate, and island milkweed bugs. *Evolution of insect migration and diapause* (pp. 254–276). Springer. https://doi.org/10.1007/978-1-4615-6941-1_11
- Dingle, H. 1996. *Migration: The biology of life on the move*. Oxford University Press.
- Dingle, H. (2001). The evolution of migratory syndromes in insects. In Woiwood, I. P., Reynolds, D. R., & Thomas, C. D. (Eds), *Insect movement: mechanisms and consequences* (pp. 159–181). Proceedings of the Royal Entomological Society's 20th Symposium, London, UK, September 1999. CABI. <https://doi.org/10.1079/9780851994567.0159>
- Dingle, H., & Drake, V. A. (2007). What is migration?. *Bioscience*, 57(2), 113–121. <https://doi.org/10.1641/B570206>
- Dupuis, J. R., & Sperling, F. A. H. (2015). Repeated reticulate evolution in North American *Papilio machaon* group swallowtail butterflies. *PLoS ONE*, 10(10), 1–26. <https://doi.org/10.1371/journal.pone.0141882>
- Dupuis, J. R., & Sperling, F. A. H. (2016). Speciation, hybridization, and conservation quandaries: What are we protecting anyway. *News of the Lepidopterists' Society*, 58(4), 202–204.
- Ellegren, H., Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, 17, 422–433. <https://doi.org/10.1038/nrg.2016.58>
- England, P. R., Osler, G. H. R., Woodworth, L. M., Montgomery, M. E., Briscoe, D. A., & Frankham, R. (2003). Effects of intense versus diffuse population bottlenecks on microsatellite genetic diversity and evolutionary potential. *Conservation Genetics*, 4(5), 595–604. <https://doi.org/10.1023/A:1025639811865>
- Fisher, R. A., Ford, E. B. (1947). The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. *Heredity* 1, 143–174. <https://doi.org/10.1038/hdy.1947.11>
- Flockhart D. T. T., Wassenaar L. I., Martin T. G., Hobson K. A., Wunder M. B., Norris D. R. (2013). Tracking multi-generational colonization of the breeding grounds by monarch butterflies in eastern North America. *Proceedings of the Royal Society B* 280, 20131087. <http://dx.doi.org/10.1098/rspb.2013.1087>
- Freedman, M. G., Dingle, H., Strauss, S. Y., & Ramírez, S. R. (2020). Two centuries of monarch butterfly collections reveal contrasting effects of range expansion and migration loss on wing traits. *Proceedings of the National Academy of Science*, 117(46), 28887–28893. <https://doi.org/10.1073/pnas.2001283117>
- Gao, B., Hedlund, J., Reynolds, D. R., Zhai, B., Hu, G., & Chapman, J. W. (2020). The 'migratory connectivity' concept, and its applicability to insect migrants. *Movement Ecology* 8(1), 1–13. <https://doi.org/10.1186/s40462-020-00235-5>
- Herrell, J., & Hazel, W. (1995). Female-limited variability in mimicry in the swallowtail butterfly *Papilio polyxenes* Fabr. *Heredity*, 75(1), 106–110. <https://doi.org/10.1038/hdy.1995.110>
- Hu, G., Chapman, J. W., Lim, K. S., Reynolds, D. R., Clark, S. J., Horvitz, N., & Sapir, N. (2016). Mass seasonal biflows of high-flying insect migrants. *Science*, 354(6319), 1584–1587. <https://doi.org/10.1126/science.aah4379>
- Hu, G., Stefanescu, C., Oliver, T. H., Roy, D. B., Brereton, T., Van Swaay, C., ... & Chapman, J. W. (2021). Environmental drivers of annual population fluctuations in a trans-Saharan insect migrant. *Proceedings of the National Academy of Sciences*, 118(26). <https://doi.org/10.1073/pnas.2102762118>
- Hung, C. M., Shaner, P. J. L., Zink R. M., Liu W. C., Chu, T. C., Huang, W. S., & Li, S. H. (2014). Drastic population fluctuations explain the rapid extinction of the Passenger Pigeon. *Proceedings of the National Academy of Sciences of the United States of America* 111(29), 10636–41. <https://doi.org/10.1073/pnas.1401526111>
- Jeffries, M. J., & Lawton, J. H. (1984). Enemy free space and the structure of ecological communities. *Biological journal of the Linnean Society*, 23(4), 269–286. <https://doi.org/10.1111/j.1095-8312.1984.tb00145.x>

- Jonzén, N., Knudsen, E., Holt, R. D., & Sæther, B. E. (2011). Uncertainty and predictability: the niches of migrants and nomads. In Milner-Gulland, E. J., Fryxell, J. M., & Sinclair, A. R. E. (Eds), *Animal migration: a synthesis* (pp. 91–109). Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199568994.001.0001>
- Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., ... & Itoh, T. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Research*, 24(8), 1384–1395.
<https://doi.org/10.1101/gr.170720.113>
- Kañuch, P., Berggren, A., & Cassel-Lundhagen, A. (2020). A clue to invasion success: genetic diversity quickly rebounds after introduction bottlenecks. *Biological Invasions*.
<https://doi.org/10.1007/s10530-020-02426-y>
- Keightley, P. D., Pinharanda, A., Ness, R. W., Simpson, F., Dasmahapatra, K. K., Mallet, J., ... & Jiggins, C. D. (2015). Estimation of the spontaneous mutation rate in *Heliconius melpomene*. *Molecular Biology and Evolution*, 32(1), 239–243.
<https://doi.org/10.1093/molbev/msu302>
- Kimura, M. (1983). The neutral theory of molecular evolution. *Cambridge University Press*. <https://doi.org/10.1017/CBO9780511623486>
- Leblois, R., Rousset, F., & Estoup, A. (2004). Influence of spatial and temporal heterogeneities on the estimation of demographic parameters in a continuous population using individual microsatellite data. *Genetics*, 166, 1081–1092. <https://doi.org/10.1534/genetics.166.2.1081>
- Leblois, R., Estoup, A., & Streiff, R. (2006). Genetics of recent habitat contraction and reduction in population size: Does isolation by distance matter? *Molecular Ecology*, 15, 3601–3615.
<https://doi.org/10.1111/j.1365-294X.2006.03046.x>
- Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., Venkat, A., Andolfatto, & P., Przeworski, M. (2012). Revisiting an old riddle: What determines genetic diversity levels within species? *PLoS Biology*, 10(9), e1001388.
<https://doi.org/10.1371/journal.pbio.1001388>
- Li, F., Zhao, X., Li, M., He, K., Huang, C., Zhou, Y., ... & Walters, J. R. (2019). Insect genomes: Progress and challenges. *Insect Molecular Biology*, 28(6), 739–758. <https://doi.org/10.1111/imb.12599>
- Li, H., & Durbin, R. (2009). Making the leap: Maq to BWA. *Mass Genomics*, 25(14), 1754–1760.
<https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26(5), 589–595.
<https://doi.org/10.1093/bioinformatics/btp698>
- Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493–496.
<https://doi.org/10.1038/nature10231>
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21): 2987–2993.
<https://doi.org/10.1093/bioinformatics/btr509>
- Lohse, K., Wright, C., Talavera, G., García-Berro, A., & Darwin Tree of Life Consortium. (2021). The genome sequence of the painted lady, *Vanessa cardui* Linnaeus 1758. *Wellcome Open Research*, 6(324), 324.
<https://doi.org/10.12688/wellcomeopenres.17358.1>
- López-Mañas, R., Pascual-Díaz, J. P., García-Berro, A., Bahleman, F., Reich, M. S., Pokorny, L., ... & Talavera, G. (2022). Erratic spatiotemporal vegetation growth anomalies drive population outbreaks in a trans-Saharan insect migrant. *Proceedings of the National Academy of Sciences*, 119(19), e2121249119.
<https://doi.org/10.1073/pnas.2121249119>
- Lyons, J. I., Pierce, A. A., Barribeau, S. M., Sternberg, E. D., Mongue, A. J., & De Roode, J. C. (2012). Lack of genetic differentiation between monarch butterflies with divergent migration destinations. *Molecular Ecology*, 21(14), 3433–3444.
<https://doi.org/10.1111/j.1365-294X.2012.05613.x>
- Mackintosh, A., Laetsch, D. R., Hayward, A., Waterfall, M., Vila, R., & Lohse, K. (2019). The determinants of genetic diversity in butterflies – Lewontin’s paradox revisited. *BioRxiv*, 1–25.
<https://doi.org/10.1101/534123>
- Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27(6), 764–770. <https://doi.org/10.1093/bioinformatics/btr011>
- Mavárez, J., Salazar, C. A., Bermingham, E., Salcedo, C., Jiggins, C. D., & Linares, M. (2006). Speciation by hybridization in *Heliconius* butterflies. *Nature*, 441(7095), 868–871. <https://doi.org/10.1038/nature04738>
- Menchetti, M., Guéguen, M., & Talavera, G. (2019). Spatio-temporal ecological niche modelling of multigenerational insect migrations. *Proceedings of the Royal Society B*, 286(1910), 20191583.
<https://doi.org/10.1098/rspb.2019.1583>
- Møller, A. P. (2008). Flight distance and population trends in European breeding birds. *Behavioral Ecology*, 19(6), 1095–1102.
<https://doi.org/10.1093/beheco/arn103>
- Møller, A. P., Rubolini, D., & Lehikoinen, E. (2008). Populations of migratory bird species that did not show a phenological response to climate change are declining. *Proceedings of the National Academy of Sciences*, 105(42), 16195–16200.
<https://doi.org/10.1073/pnas.0803825105>
- Motro, U., & Thomson, G. (1982). On heterozygosity and the effective size of populations subject to size changes. *Evolution*, 36, 1059–1066.
<https://doi.org/10.2307/2408083>
- Murray, G. G. R., Soares, A. E. R., Novak B. J., Schaefer N. K., Cahill, J. A., Baker, A. J., Demboski, J. R., ... & Gilbert, M. T. P. (2017). Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science*, 358(6365), 951–954.
<https://doi.org/10.1126/science.aao0960>
- Nadachowska-Brzyska, K., Burri, R., Smeds, L., & Ellegren, H. (2016). PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Molecular Ecology*, 25(5), 1058–1072. <https://doi.org/10.1111/mec.13540>
- Nevo, E. (1978). Genetic variation in natural populations: Patterns and theory. *Theoretical Population Biology*, 13(1), 121–177.
[https://doi.org/10.1016/0040-5809\(78\)90039-4](https://doi.org/10.1016/0040-5809(78)90039-4)
- Newton, I. 2008. *The ecology of bird migration*. Academic Press.
- Nishimura, O., Hara, Y., & Kuraku, S. (2017). GVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics*, 33(22), 3635–3637.
<https://doi.org/10.1093/bioinformatics/btx445>
- Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2016). Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, 32(2), 292–294.
<https://doi.org/10.1093/bioinformatics/btv566>
- Owen, D. F. (1987). Winter breeding by *Cynthia cardui* (L.) (Lepidoptera: Nymphalidae) in Crete and Madeira, and the possible significance of parasitoids in initiating migration. *Entomologist’s Gazette*, 38, 11–12.
- Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401, 877–884. <https://doi.org/10.1038/44766>
- Palstra, F. P., & Ruzzante, D. E. (2008). Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence?. *Molecular ecology*, 17(15), 3428–3447.
<https://doi.org/10.1111/j.1365-294X.2008.03842.x>
- Patil, A. B., & Vijay, N. (2021). Repetitive genomic regions and the inference of demographic history. *Heredity*, 127(2), 151–166.
<https://doi.org/10.1038/s41437-021-00443-8>
- Pfeiler, E., & Markow, T. A. (2017). Population connectivity and genetic diversity in long-distance migrating insects: divergent patterns in representative butterflies and dragonflies. *Biological Journal of the Linnean Society*, 122(2), 479–86.
<https://doi.org/10.1093/biolinnea/blx074>

- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., & Van Willigen, B. (2016). Package 'nlme'. Linear and nonlinear mixed effects models. *R package version*, 3(57), 1–89.
- Pollard, E., van Swaay C. A. M., Stefanescu C., Lundsten, K. E., Maes, D., & Greateorex-Davies, J. N. (1998). Migration of the painted lady butterfly *Cynthia cardui* in Europe: Evidence from monitoring. *Diversity and Distributions*, 4, 243–253. <https://www.jstor.org/stable/2999830>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Raymond, L., Plantegenest, M., Gauffre, B., Sarthou, J. P., & Vialatte, A. (2013). Lack of genetic differentiation between contrasted overwintering strategies of a major pest predator *Episyrphus balteatus* (Diptera: Syrphidae): implications for biocontrol. *PLoS one*, 8(9), e72997. <https://doi.org/10.1371/journal.pone.0072997>
- Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in ecology and evolution*, 3(2), 217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>
- Ryan, S. F., Lombaert, E., Espeset, A., Vila, R., Talavera, G., Dinca, V., ... & Shoemaker, D. W. (2019). Global invasion history of the agricultural pest butterfly *Pieris rapae* revealed with genomics and citizen science. *Proceedings of the National Academy of Sciences of the United States of America*, 116(40), 20015–20024. <https://doi.org/10.1073/pnas.1907492116>
- Sanderson, J. F., Donald F. P., Pain J. D., Burfield J. I., & van Bommel P. J. F. (2006). Long-term population declines in Afro-Paleartic migrant birds. *Biological Conservation*, 131(1), 93–105. <https://doi.org/10.1016/j.biocon.2006.02.008>
- Saunders, S. P., Ries, L., Neupane, N., Ramírez, M. I., García-Serrano, E., Rendón-Salinas, E., & Zipkin, E. F., (2019). Multiscale seasonal factors drive the size of winter monarch colonies. *Proceedings of the National Academy of Sciences of the United States of America* 116(17), 8609–14. <https://doi.org/10.1073/pnas.1805114116>
- Satterfield, D. A., Sillett, T. S., Chapman, J. W., Altizer, S., & Marra, P. P. (2020). Seasonal insect migrations: massive, influential, and overlooked. *Frontiers in Ecology and the Environment*, 1–10. <https://doi.org/10.1002/fee.2217>
- Shields, O. (1992). World distribution of the *Vanessa cardui* group (Nymphalidae). *Journal of the Lepidopterists' Society*, 46(3), 235–238.
- Shipilina, D., Näsvall, K., Höök, L., Vila, R., Talavera, G., & Backström, N. (2022). Linkage mapping and genome annotation give novel insights into gene family expansions and regional recombination rate variation in the painted lady (*Vanessa cardui*) butterfly. bioRxiv. <https://doi.org/10.1101/2022.04.14.488360>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Stefanescu, C., Askew, R. R., Corbera, J., & Shaw, M. R. (2012). Parasitism and migration in southern Palaearctic populations of the painted lady butterfly, *Vanessa cardui* (Lepidoptera: Nymphalidae). *European Journal of Entomology*, 109(1), 85–94. <https://doi.org/10.14411/eje.2012.011>
- Stefanescu, C., Páramo, F., Åkesson, S., Alarcón, M., Ávila, A., Brereton, T., ... & Chapman, J. W. (2013). Multi-generational long-distance migration of insects: studying the painted lady butterfly in the Western Palaearctic. *Ecography*, 36(4), 474–486. <https://doi.org/10.1111/j.1600-0587.2012.07738.x>
- Talavera, G., & Vila, R. (2017). Discovery of mass migration and breeding of the painted lady butterfly *Vanessa cardui* in the Sub-Sahara: The Europe-Africa migration revisited. *Biological Journal of the Linnean Society*, 120(2), 274–285. <https://doi.org/10.1111/bij.12873>
- Talavera, G., Bataille, C., Benyamini, D., Gascoigne-Pees, M., & Vila, R. (2018). Round-trip across the Sahara: Afrotropical Painted Lady butterflies recolonize the Mediterranean in early spring. *Biology Letters*, 14(6), <https://doi.org/10.1098/rsbl.2018.0274>
- Talla, V., Soler, L., Kawahami, T., Dincă V., Vila, R., Friberg, M., Wiklund, C., Backström, N. (2019). Dissecting the effects of selection and mutation on genetic diversity in three wood white (*Leptidea*) butterfly species. *Genome Biology and Evolution*, 11(10), 2875–2886. <https://doi.org/10.1093/gbe/evz212>
- Talla, V., Pierce, A. A., Adams, K. L., de Man, T. J. B., Nallu, S., Villablanca, F. X., Kronforst, M. R., & de Roode, J. C. (2020). Genomic evidence for gene flow between monarchs with divergent migratory phenotypes and flight performance. *Molecular Ecology*, 29(14), 2567–2582. <https://doi.org/10.1111/mec.15508>
- Vandenbosch, R. (2003). Fluctuations of *Vanessa cardui* butterfly abundance with El Niño and Pacific Decadal Oscillation climatic variables. *Global Change Biology*, 9, 785–790. <https://doi.org/10.1046/j.1365-2486.2003.00621.x>
- Vucetich, J. A., Waite, T. A., & Nunney, L. (1997). Fluctuating population size and the ratio of effective to census population size. *Evolution*, 51: 2017–2021. <https://doi.org/10.2307/2411022>
- Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics*, 33(14), 2202–2204. <https://doi.org/10.1093/bioinformatics/btx153>
- Wang, J. (2005). Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society B*, 360, 1395–1409. <https://doi.org/10.1098/rstb.2005.1682>
- Webster, M. S., Marra, P. P., Haig, S. M., Bensch, S., & Holmes, R. T. (2002). Links between worlds: unraveling migratory connectivity. *Trends in ecology & evolution*, 17(2), 76–83. [https://doi.org/10.1016/S0169-5347\(01\)02380-1](https://doi.org/10.1016/S0169-5347(01)02380-1)
- White, E. P., Ernest, S. M., Kerkhoff, A. J., & Enquist, B. J. (2007). Relationships between body size and abundance in ecology. *Trends in Ecology & Evolution*, 22(6), 323–330. <https://doi.org/10.1016/j.tree.2007.03.007>
- Whitlock, M. (1992). Temporal fluctuations in demographic parameters and the genetic variance among populations. *Evolution*, 46, 608–615. <https://doi.org/10.1111/j.1558-5646.1992.tb02069.x>
- Wilcove, D. S., & Wikelski, M. (2008). Going, going, gone: Is animal migration disappearing? *PLoS Biology*, 6(7), 1361–1364. <https://doi.org/10.1371/journal.pbio.0060188>
- Williams, C. B. (1970). The migrations of the painted lady butterfly, *Vanessa cardui* (Nymphalidae), with special reference to North America. *Journal of the Lepidopterists' Society*, 24, 157–175.
- Wilson, K., & Gatehouse, A. G. (1993). Seasonal and geographical variation in the migratory potential of outbreak populations of the African armyworm moth, *Spodoptera exempta*. *Journal of Animal Ecology*, 62, 169–181. <https://doi.org/10.2307/5491>
- Wilson, S., Saracco, J. F., Krikun, R., Flockhart, D. T. T., Godwin, C. M., Foster, K. R. (2018). Drivers of demographic decline across the annual cycle of a threatened migratory bird. *Scientific Reports*, 8:7316. <https://doi.org/10.1038/s41598-018-25633-z>
- Wright, S. (1948). On the roles of directed and random changes in gene frequency in the genetics of populations. *Evolution* 2, 279–294. <https://doi.org/10.1111/j.1558-5646.1948.tb02746.x>
- Zhan, S., Zhang, W., Niitepöld, K., Hsu, J., Haeger, J. F., Zalucki, M.P., ... Kronforst, M. R. (2014). The genetics of monarch butterfly migration and warning coloration. *Nature*, 514(7522), 317–321. <https://doi.org/10.1038/nature13812>