Title: Population genetics of a recent range expansion and subsequent loss of migration in monarch butterflies

Running Title: Pacific monarch butterfly genetics

Authors: William B. Hemstrom1\*, Micah G. Freedman2,3\*, Myron P. Zalucki4, Santiago R. Ramírez2,3, Michael R. Miller1

\*Authors contributed equally

Author Affiliations:

1. Department of Animal Science, University of California, Davis

2. Department of Evolution and Ecology, University of California, Davis

3. Center for Population Biology, University of California, Davis

4. School of Biological Sciences, The University of Queensland, Australia, 4072

**Abstract**

Range expansions—whether permanent or transient—strongly influence the distribution of population genetic variation in space. Monarch butterflies are best-known for their long-distance migration in North America but have greatly expanded their geographic range over recent evolutionary history, including an expansion across the Pacific Ocean. Recently-established monarch populations generally form non-migratory, year-round breeding populations. Previous research has highlighted the stepwise nature of this range expansion, though questions remain about its timing and the population genetic consequences of migration loss. Here, we present reduced-representation sequencing data for 281 monarchs from North America and 15 locations across the Pacific, with the goal of understanding (1) how the monarch’s expansion across the Pacific has broadly shaped patterns of population genetic variation and (2) how loss of migration has influenced fine-scale spatial patterns of differentiation. We find support for previously described stepwise dispersal across the Pacific, but also document an additional westward expansion from Hawaii into the Mariana Islands. Monarchs within the Mariana Islands show strong patterns of differentiation, despite being in extremely close proximity; by contrast, migratory North American samples form a single genetically panmictic population across the entire continent. Estimates for the timing of the monarch’s establishment in the Pacific have a high uncertainty (approximately 100 to 1,000,000 years ago), but do not exclude the recent expansions supported by historical records. . Together, our data support (1) a single recent westward expansion across the Pacific whose timing accords with available historical records of establishment and (2) a strong role for seasonal migration in determining patterns of spatial genetic variation.

**Key words:** Range expansion, serial dispersal, monarch butterfly, population genomics

**Introduction**

Species that undergo range expansions often show distinctive signatures of population genetic variation in space. Over extended time scales, geographic range expansions generally involve decreasing relatedness and increasing contributions of genetic drift in populations further from the original source population (Hewitt 1996, Excoffier et al. 2009). This pattern is evident in serial stepwise expansion events, in which populations are founded in a stepping-stone fashion (Ibrahim et al. 1996, Slatkin and Excoffier 2012). Serial dispersal is characteristic of many post-glacial range expansions into temperate regions and has been shown for species including eider ducks (*Somateria mollissima*) (Tiedemann et al. 2004), ragwort (*Senecio helleri*) (Bettin et al. 2007), and rough-skinned newts (*Taricha granulosa*) (Kuchta and Tan 2005).

Studies on the population genetics of geographic range expansions tend to focus on expansion events that occur over extended time scales, though the same expansion processes characterize annual movements associated with seasonal migration. Unlike permanent range expansion, however, seasonal migration may involve individuals capable of making round-trip journeys and traversing the entire species range in their lifetime (Dingle et al. 2014), thereby limiting opportunities for genetic divergence in allopatry. In species that migrate seasonally, patterns of population genetic variation in space are best captured by considering migratory connectivity of breeding populations (e.g. Cohen et al. 2018, Gao et al. 2020). For example, Wilson’s warbler (*Wilsonia pusilla*) has eastern and western North American summer breeding populations that are genetically distinct, despite sharing an overwintering range in Central America (Irwin et al. 2011), and anadromous Coho (*Oncorhynchus tshawytscha*) and steelhead (*O. mykiss*) salmon populations show strong genetic differentiation corresponding to the river drainages where they spawn (Waples et al. 2004, Prince et al. 2017). By contrast, Japanese eels (*Anguilla japonica*) that share common spawning grounds but migrate to disparate areas across temperate Asia show little evidence for genetic differentiation over time or space (Gong et al. 2019).

Among migratory species, population-level differentiation is often most pronounced in species that show evidence for partial migration, whereby species are comprised of both migratory and non-migratory populations (Chapman et al. 2011). This phenomenon is common across the tree of life and has been documented in birds (Adriaensen and Dhondt 1990), insects (Menz et al. 2019), and ungulates (Berg et al. 2019). Although the evolutionary origins of partial migration are sometimes unclear, one recently invoked scenario involves a migratory, geographically widespread lineage giving rise to one or more non-migratory descendent lineages that become genetically distinct due to mismatches in the timing and/or location of breeding. This scenario has been hypothesized to be an important contributor to patterns of speciation in tropical birds (e.g. Kondo et al. 2008, Gomez-Bahamon et al. 2020).

Monarch butterflies (*Danaus plexippus* (L.)) provide an intriguing system for studying the effects of both global range expansion and loss of migration on spatial population genetic structure. Migratory North American monarchs comprise a single genetically indistinguishable population (Lyons et al. 2012, Talla et al. 2020), despite having a summer breeding range that covers most of the North American continent. Over recent evolutionary history, monarchs have expanded their range globally (Ackery and Vane-Wright 1984, Vane-Wright 1993, Zalucki and Clarke 2004, Zhan et al. 2014, Fernández-Haeger et al. 2015, Pierce et al. 2015), with a southern expansion into South America and the Caribbean, an eastward expansion across the Atlantic and into the Iberian Peninsula, and a westward expansion across the Pacific. In contrast to their migratory North American ancestors, these expansion populations generally form non-migratory, year-round breeding populations in areas where they become established (Zhan et al. 2014). The lone exception to this pattern is in southern Australia and New Zealand, where monarchs move seasonally and form overwintering clusters akin to those observed in western North America (Wise 1980, James 1993). In this paper, we focus exclusively on the monarch’s expansion and subsequent loss of migration in locations across the Pacific.

Little is currently known about how contemporary loss of migration has affected fine-scale patterns of population differentiation in monarchs. Two studies have addressed this question: Hughes and Zalucki (1984) used four allozyme markers and found relatively high FST between host plant patches over small spatial scales (tens to hundreds of kilometers) in Queensland, Australia. Pierce et al. (2014b) used microsatellite markers and showed that monarchs from the Hawaiian archipelago show little differentiation among islands. However, the conclusions of these studies are limited by the spatial scale of sampling and the number of loci studied. Furthermore, the timing of the monarch’s Pacific expansion remains uncertain. Demographic simulations indicate that establishment timing may have happened as long as 2000-3000 years ago (Zhan et al. 2014), although these estimates conflict with historical records, which suggest that expansion across the Pacific happened between the years 1840-1900 (Zalucki and Clarke 2004, Freedman et al. 2020a).

In this study, we sequenced 281 monarch butterflies at >70,000 highly variable genomic sites from the ancestral North American population and many Pacific Island populations, including a number of previously unsampled locations: the Mariana Islands (Guam, Rota, and Saipan) and Norfolk Island. The goals of this study were to understand (1) overall patterns of relatedness among Pacific and North American populations; (2) the timing of range expansion and the amount of ongoing gene flow between North American and Pacific locations; (3) how migratory and non-migratory populations differ in their distribution of population genetic variation in space.

**Methods**

*Sample preparation and sequencing*

Monarchs were collected as either larvae or adult butterflies from locations around the world between 1990-2017 (Figure 1a and Table S1).DNA was extracted from samples using a magnetic bead-based protocol (Ali et al. 2016) and quantified using Quant-iT PicoGreen dsDNA Reagent (Thermo Fisher Scientific) on a FLx800 Fluorescence Reader (BioTek Instruments). Restriction Associated Digest (RAD) DNA libraries were then created using the PstI restriction enzyme according to Ali et al. (2016) and sequenced using 150 bp paired-end sequencing on an Illumina Hi-Seq 4000.

*Sequence alignment, filtering, and genotype calling*

We aligned raw sequence data to version 3 of the monarch butterfly genome assembly [(Zhan](https://paperpile.com/c/tNxuHC/psiG) and Reppert 2013) using the mem algorithm implemented in Burrows-Wheeler Aligner [(Li](https://paperpile.com/c/tNxuHC/7NQQ) and Durbin 2009). Sequence data was sorted and filtered for PCR duplicates and improper pairs using SAMtools [(](https://paperpile.com/c/tNxuHC/EXik)Li et al. 2009). For use in demographic reconstruction, genotypes were then called using the SAMtools genotype likelihood model [(Li](https://paperpile.com/c/tNxuHC/EXik) et al. 2009) as implemented in the ANGSD software package with a minimum mapping and base call quality score of 20, a SNP *p*-value of 1e-8, a uniform genotype prior, and a posterior genotype probability cutoff of 0.95 [(Korne](https://paperpile.com/c/tNxuHC/hYyB)liussen et al. 2014). For analyses that relied on called genotypes, we removed potential paralogous sites using the ngsParalog tool by removing all sites within a kilobase of any SNP with a log ratio test statistic of > 10 in any population [(](https://paperpile.com/c/tNxuHC/OQZa)Linderoth 2018). Individuals with less than 75% of called SNPs were removed. Since strong bottlenecks are likely to remove rare SNPs from the population, we did not use a minor allele frequency filter when calling genotypes for the calculation of the basic diversity statistics (π, HO, and Het/Hom ratio).

*Patterns of relatedness among monarch populations*

We calculated the average number of pairwise differences (µ), observed heterozygosity (Ho), heterozygote/homozygote ratio per individual (Het/Hom), Tajima’s D, and fixation index (FST) between each pair of populations for each SNP using the snpR package (Hemstrom and Jones 2021). We calculated FST using the R implementation of the GENEPOP software package [(Rousset](https://paperpile.com/c/tNxuHC/UGgy) 2008) with a minor allele frequency cutoff of 0.05. To calculate Tajima’s D (Tajima 1989), we used all sequenced sites that passed the quality and paralog filters without removing non-paralogous sites or those with low minor allele frequencies. For each of these statistics, the eastern and western North American samples were pooled together.

In order to describe basic population structure, we created a neighbor-joining tree (Saitou and Nei 1987) using the ape R package v.5.0 [(](https://paperpile.com/c/tNxuHC/jHGn)Paradis and Schliep 2019). In order to maximize the amount of genetic data contributing to this tree, the input distance matrix was created using the Identity-by-State approach in ANGSD with the same parameters as above, save for a minor allele frequency cutoff of 0.05 (Korneliussen et al. 2014). A Principal Component Analysis (PCA) was also conducted using this dataset. For comparison, NGSadmix was used to generate individual ancestry coefficients for each individual for between 1 and 9 putative population clusters (k) (Skotte et al. 2013). Each value of k was run 10 times, and the results were collapsed into consensus plots using CLUMPP (Jakobsson and Rosenberg 2007). The pophelper (Francis 2017) and snpR (Hemstrom et al. *in prep*) R packages were used to simplify these analyses. No individuals were removed for this analysis.

To quantify the direction and strength of population spread across the Pacific, we calculated the directionality index (***ψ***) (Peter and Slatkin 2013) for each pairwise combination of North America, Hawaii, Queensland, Guam, Rota, Norfolk Island populations using the snpR package (Hemstrom et al *in prep*). We created the polarized site-frequency spectra used in these calculations using the δaδi (dadi) (Charles et al. 2018) dataset described below by projecting populations down to ten gene copies each using the methods described by Gutenkunst et al. (2009) as implemented in snpR (Hemstromand Jones 2021). Using 10 gene copies tended to produce the highest number of maintained SNPs in the resulting spectra. The SNPs were polarized via reference to whole genome sequence data of the best-sequenced individual of the monarch sister taxon *Danaus erippus* (Zhan et al. 2014) by alignment to the monarch genome as described above.

In order to determine the effect of migration on population connectivity, we looked for evidence of isolation by distance (IBD) between samples from the Mariana Islands, North America, Australia, and Hawaii. To do so, we first filtered out all loci with a minor allele frequency of less than 0.05 in the population being examined, calculated Edwards’ angular genetic distance (Edwards 1971) between each pair of samples from the given populations, and then compared these distances to the geographic distances between samples using a Mantel Test (Mantel 1967).

*Demographic history of the monarch’s expansion*

To describe the patterns of establishment and migration between North America and the Pacific, the demographic reconstruction program dadi (Gutenkunst et al. 2009) was used to estimate the demographic history of the North American and Hawaiian samples. In order to reduce potential bias due to linkage, filtered SNPs were randomly subsampled such that no locus was within 10,000 bp of another using a custom R script. The resulting 11,384 SNPs were then projected down to a sample size of 100 gene copies from North America and 10 from Hawaii, resulting in 9370 total SNPs. These projection numbers were picked to maximize the remaining number of SNPs in the dataset.

We fit a range of possible models to the observed data: (1) each of the “Island Model” set described in the dadi\_pipeline (Portik et al. 2017), which contains some models originally published in Charles et al. (2018); (2) the model described by Zhan et al. (2014) for the same comparison; (3) a similar model that allowed for an additional period of growth prior to the establishment of the Hawaiian population and another following establishment (hereafter referred to as the *Three Epoch* model). The latter model is shown in greater detail in Figure S1. Each of the dadi\_pipeline models and their logistic growth versions were run three times: once with growth allowed in the founding population post-split, once with growth allowed in the founded population post-split, and once with growth allowed in both populations post-split. Note that in each of these (dadi\_pipeline) models, a source population splits to form two descendant populations, with an optimized parameter (s) controlling the portion of the population that forms each descendant population. When s is optimized to be very small (as it typically was), the founded population represents only a very small proportion of the ancestral population, as is likely realistic for the founding of the Hawaiian population from the North American population. Graphical depictions of the dadi\_pipeline models are available in Portik et al. (2017) and Charles et al. (2018).

To optimize the models we fit during the analysis, we used a variation on dadi\_pipeline, the sequential step-down parameter permutation approach described by Portik et al. (2017). Unlike this method however, we set the starting parameters for each sequential run via weighting the parameters from each run in the previous iteration by the relative AIC score of that iteration, such that all but the worst runs contribute in some degree to the starting parameters for the next step. The number of runs and iterations per step are listed in Table S3. Individual optimization runs were killed if they took longer than 48hrs to complete; these runs tended to take far longer to finish and often included integration errors due to extremely small population sizes resulting in extremely large amounts of genetic drift. Most runs completed in under 48hrs and are included in the results (Table S5).

To extract meaningful parameter units from the results, we assumed 0.3 years per generation and used the per-base mutation rate of 8.4x10-9 reported from Drosophila melanogaster (Haag-Liautard et al. 2007). We use these values to match those used by Zhan et al. (2014) for ease of comparison. Using a potentially more realistic generation time of 7 generations per year resulted in more recent divergence times, and using the slower mutation rate reported for the more closely related *Heliconius melpomene* of 2.9x10-9 (Keightley et al. 2015) resulted in larger effective size estimates and more distant divergence times, for a net result of slightly more distant divergence times and larger effective sizes. Overall, the results did not differ qualitatively to any substantial degree. In order to determine the length of the considered genomic region, we multiplied the total number of bases sequenced after quality filtering (but not SNP p-value filtering so as to count non-polymorphic sites) by the ratio of SNPs in the final allele frequency spectrum to the total number of called SNPs.

Among the large set of possible demographic models, the simple *Found and Grow* scenario (Figure S1a), which had a constant ancestral population size in North America, Hawaii colonization, and then population growth in both sites, produced the lowest AIC scores on the final pass of the pipeline. However, the *Two Epoch* model, which had a single admixture event but no consistent migration (Figure S1b), had the lowest AIC score across all passes of the pipeline. Thenew *Three Epoch* , which involved multiple rounds of demographic expansion in the ancestral North American population, followed by colonization and growth in Hawaii, had a lower AIC score than *Found and Grow* across all all passes and a lower AIC score than the *Two Epoch* on the final pass(Figure S2)*.* These three models were top three models in across all passes and in the final pass alone(Figure S2, Table S5). The *Three Epoch* model is a more complex version of the model specified in Zhan et al. (2014, hereafter *Zhan*). We report in detail the results of the *Found and Grow*, *Three Epoch, Two Epoch, and Zhan* models and, where relevant, highlight discrepancies in the inferences that they produce.

**Results**

*Sequencing results*

After paralog filtering, we were able to genotype 2,159,978 sites in at least 50% of individuals. 541,899 of these sites were polymorphic, and 71,157 had a minor allele frequency above 0.05. After removing individuals sequenced at less than 75% of called genotypes, 413,271 and 70,878 sites remained in each category, respectively. The number of samples from each population after filtering can be found in Table 1.

*Overall patterns of relatedness*

Principal component analysis separated North American, Hawaiian, Mariana Islands, and southwest Pacific samples along two axis of expansion (Figure 1a). Directionality index (***ψ***) scores generally indicated westward establishments (Figure S3). Genetic diversity (π, HO, and Het/Hom), was highest in the ancestral North American populations, followed by Hawaii, Australia, and then the remaining Pacific Island populations (Table 1, Figure S4). Likewise, Tajima’s D is positive in all sites besides North America and Hawaii (Table 1). FST and NGS results both reflect the patterns we observed in the PCA (Table S2, Figure 1b).

*Patterns of differentiation within expansion populations*

Samples from the Mariana Islands (especially the well-sampled Guam and Rota populations) appear to form highly distinct populations, despite their close physical proximity (Figure 1a-c). By contrast, populations within Hawaii (Maui and Oahu) and Australia (Queensland, New South Wales, and Victoria) do not show strong patterns of differentiation (Figure 1a-c). Norfolk Island, the other previously unsampled population in our dataset, groups closely with samples from Australia and New Zealand (Figure 1a-c). IBD patterns were strongest between the Mariana island samples (p = 0.001), were present but not significant in Hawaii and Australia (p = 0.086 and 0.0766, respectively), and were absent in North America (p = 0.489).

*Timing of establishment and patterns of ongoing gene flow*

Our three best-performing demographic models (Found and Grow*, Three Epoch*, and *Two Epoch*) and the *Zhan* model gave highly variable estimates of establishment timing and founding populations size, with the *Three Epoch* model generally producing much broader estimates for these parameters. For example, the simpler *Found and Grow*, *Two Epoch*, and *Zhan* models were consistent in indicating a founding time of approximately 104-105 years ago, while the *Three Epoch* model suggested establishment times that ranged between approximately 102 to 105 years ago (Figure 2). Similarly, the *Found and Grow*, *Two Epoch*, and *Zhan* models were more consistent in predicting a large founding population of 103-106 individuals, while the *Three Epoch* models suggested a broader founding population size of between 10 and 106 individuals (Figure 2). These models also differed in their estimates of the contemporary *Ne* for the Hawaiian population, with the *Found and Grow* and *Two Epoch* model suggesting a large *Ne* of around 106 and the *Three Epoch* and *Zhan* models generally producing estimates of Hawaiian *Ne* between 102 and 107 (Figure 2). For all of the models, the major discrepancy between the observed and simulated site frequency spectra tended to be that the models underestimated the number of rare derived alleles in the North American but not the Hawaiian populations (Figure 3, Figure S5). This may be indicative that the models did not optimize for strong enough founder effects in Hawaii, which would have caused a more drastic loss of rare alleles. This may be due to the fact that processing times for dadi increase drastically when populations sizes are very small, as do the risk for integration errors. Since we omitted runs with these issues, we could not move as readily into these areas of the parameter space during our model optimizations.

For other parameters, the *Found and Grow, Two Epoch,* and *Three Epoch and grow* models generated similar estimates. Both models suggest very low levels of contemporary migration between North America and Hawaii, with the *Found and Grow* and *Two Epoch* models converging near 0 for both directions and the *Three Epoch and grow* models generally suggesting migration rates of < 5 x 10-4 and < 2.5 x10-4 for individuals per generation from North America to Hawaii and from Hawaii to North America, respectively, and the Zhan model suggesting migration rates of < 2.5 x 10-4 in either direction (Figure 2). This finding accords with our intuition that trans-oceanic dispersal events in monarchs should be rare.

**Discussion**

Our analysis suggests a recent natural range expansion characterized by serial stepwise dispersal across the Pacific in monarchs. We find evidence for two independent expansions upon establishment in Hawaii, with a previously uncharacterized westward expansion from Hawaii into the Mariana Islands. Our overall results provide a higher resolution picture of the monarch’s pattern of establishment in the Pacific and show that historical estimates of introduction timing overlap with demographic reconstructions. Lastly, we find that loss of seasonal migration is accompanied by increased levels of isolation by distance.

Many geographic range expansions occur via serial stepwise dispersal, and we found strong evidence for this pattern in Pacific monarchs, consistent with a previous study (Pierce et al. 2014a). Monarchs in the Mariana Islands are the product of a distinct expansion event within the Pacific. Summary statistics support a scenario of directional dispersal from North America to Hawaii, from Hawaii to Guam, and from Hawaii to Australia. This pattern is reflected in both the positive directionality index measures (0.07, 0.08, and 0.05, respectively) (Peter and Slatkin 2013) and other summary statistics, such as the general increase in Tajima’s D across the Pacific, consistent with stronger or more recent population bottlenecks during successive colonization. Interestingly, monarch populations in Hawaii and Australia seem to maintain relatively high levels of genetic diversity, despite the apparent bottlenecks associated with establishment. This is especially striking in the Australian population, which was itself likely founded by individuals from a much smaller population in New Caledonia (Clarke and Zalucki 2004). The retention of genetic diversity in Hawaii and Australia may reflect rapid population growth upon establishment, which could temper the loss of allelic diversity that might be predicted with a bottleneck event, akin to the scenario described in Hawaiian *Drosophila* by Nei et al. (1975). The slightly negative Tajima’s D value in Hawaii is consistent with population growth following a bottleneck, and is thus consistent with this hypothesis.

Within the Mariana Islands, there was a strong pattern of differentiation between islands, especially between the nearby islands of Guam and Rota. This pattern is striking because of their close geographic proximity: these islands are separated by only 40 km of open ocean. By contrast, our samples from North America, despite coming from overwintering sites nearly 2000 km apart, formed a single genetically indistinguishable population. This pattern is apparent from the strong pattern of IBD observed within Mariana Islands samples compared to weak/no IBD within North America. Our results are similar to those of Dapporto et al. (2017), who also noted strong genetic differentiation between butterfly populations even from nearby islands, as well as Alvial et al. (2018), who showed that a migratory dragonfly exhibits little genetic differentiation across its migratory Central and South American range but substantial genetic differentiation between non-migratory populations on Pacific islands.

The lack of differentiation within North American monarchs is consistent with other population genetic analyses of eastern and western North American monarchs (Brower and Boyce 1991, Shephard et al. 2002, Lyons et al. 2012, Zhan et al. 2014, Talla et al. 2020) and is consistent with studies that have suggested movement of individuals between eastern and western North America (Dingle et al. 2005, Morris et al. 2015, Billings 2019). The strong population genetic differentiation within the Mariana Islands but not at the scale of the entire North American continent highlights both (1) the pervasive role that long-distance migration in North America plays in collapsing any patterns of population structure that might otherwise develop, and (2) the fact that many Pacific monarch populations likely have extremely small effective population sizes that are susceptible to very strong genetic drift.

In contrast to populations within the Mariana Islands, samples from Hawaiian and Australian monarchs show only modest evidence for IBD that might be expected in non-migratory monarch populations. Within Hawaii, our samples from Maui and Oahu formed a single genetic cluster, consistent with the results of Pierce et al. (2014b). Likewise, Australian samples from New South Wales and Victoria grouped with samples from Queensland. This result differs somewhat from the results of Hughes and Zalucki (1984), who reported considerable among-site genetic variation within Queensland, but is consistent with similar later work (Zalucki et al. 1987).

For Hawaiian monarchs, it is not immediately clear why the islands of Maui and Oahu do not form clearly distinct populations. One possibility is that prevailing winds promote gene flow between islands in a way that differs from the Mariana Islands. Pacific monarchs are likely moved by wind patterns, similar to wind-driven movement patterns noted in migratory *Vanessa cardui* (Stefanescu et al. 2007), and it has been suggested that a tropical cyclone may have led to the monarch’s establishment in Australia (Clarke and Zalucki 2004). Another possibility is between-island movement of monarchs by butterfly breeders in Hawaii, who sell monarchs for release at weddings and celebrations. In the case of Australian monarchs, the lack of strong differentiation across the continent may be driven by seasonal migration patterns akin to those seen in western North American monarchs (James 1993, James and James 2019). Australian monarchs retain migration-associated behaviors that further support the notion that they may undergo large-scale seasonal movements (James 1993, Freedman et al. 2018, Hemstrom et al. *in prep*). Thus, the lack of continent-wide population structure seen in migratory North American monarchs may be recapitulated, albeit to a lesser extent, in Australia.

Interpreting the results of our demographic models is somewhat more complicated than interpreting basic patterns of relatedness among populations. This is due to the conflicting inferences provided by the two best-performing model structures and the wide range of parameter estimates in the *Three Epoch* models. Although we present the results of both the simpler *Found and Grow* and the more complicated *Three Epoch* models, we are inclined to place more confidence in the estimates produced by the *Three Epoch* models for two reasons: (1) the demographic scenario that it specifies—recent demographic expansion in the ancestral North American population prior to geographic expansion—has empirical support from other studies (Zhan et al. 2014, Pfeiler et al. 2017) and accords with our understanding of past changes in climate, and (2) this model structure produces parameter estimates that match our prior understanding for how and when monarch range expansion may have occurred. The latter point is related to the former: because a North American population expansion is not allowed until after the founding of Hawaii in the *Found and Grow* model, this model forces an ancient founding of the Hawaiian population in order to allow for the ancient growth of the North American population. As such, we focus our discussion on the estimates produced by the *Three Epoch* model.

In general, our demographic results do not exclude a recent founding of the Hawaiian population by North American monarchs (Fig 4d). While our model optimizations span several orders of magnitude for the time since establishment, many of the iterations settled on introduction estimates of less than 200 years ago for the *Three Epoch* model. Since the earliest historical records of monarchs on Hawaii date to roughly 200 years ago (1841) (Zalucki and Clarke 2004), we are inclined to accept the results of iterations with shorter estimated divergence times. Other lines of evidence supporting recent (<200 years) Hawaiian establishment include: (1) the lack of noticeable phenotypic differentiation between North American and Pacific Island monarchs, especially relative to the pronounced phenotypic differences in non-migratory populations from the Caribbean and South America (Freedman et al. 2020a), which have historically been treated as separate subspecies (Ackery and Vane-Wright 1984), (2) the likely need for human-mediated transport of the monarch’s host plants (some of which are native to subtropical Africa) as a pre-condition of monarch establishment in the Pacific, and (3) recent genomic evidence showing that captive breeding of monarchs over short time scales is sufficient to generate patterns of genetic divergence comparable to those observed between North American and Pacific populations (Tenger-Trolander et al. 2019). Interestingly, our re-implementation of the model used by Zhan et al. (2014) produced results that were similar to theirs, with the vast majority of model iterations supporting an introduction time of roughly 1000+ years ago (Figure S5). This highlights the need to run a range of possible demographic models when attempting to infer demographic history, since failing to account for underlying complexity in population histories can result in very divergent parameter estimates.

In addition, our models consistently underestimated the number of rare, derived alleles present in North America but not Hawaii. During very strong bottlenecks, we would expect many rare alleles to be lost, suggesting that our models may be overestimating the founding population size, and thus likely establishment date. Since dadi can struggle to calculate site frequency spectra when population sizes are very small due to very large amounts of drift, iterations that optimize to this segment of parameter space are more likely to have integration errors or very long processing times. These uncompleted runs were not included in our model results, and so this part of the parameter space may be inadequately explored. This also suggests that it would be unwise to rule out a recent introduction with a very strong bottleneck based on a demographic analysis alone.

Demographic model results were also variable in their estimates of founding population sizes in Hawaii. Some models produced estimates as high as 10,000 founding individuals, which seems implausible given the incredibly long distance (>3500 km) between North America and Hawaii. The extremely wide range of parameter estimates for founding population size and timing may reflect that, in practice, it is difficult to distinguish between a very recent strong bottleneck versus a more distant but less severe bottleneck. Our model results do are consistent with this, since the model iterations with recent establishments also tended to have smaller establishment population sizes (Figure 2).

In contrast to variable estimates of establishment timing and founding population size, demographic models were consistent in suggesting very low contemporary migration rates (on the order of 0.0001 individuals per generation from North America to Hawaii and vice versa). Our results thus contrast with those of Pierce et al. (2014a), who inferred much higher migration rates (nearly 10 individuals/generations) between North America and Hawaii. We are more confident in our results due to (1) the much larger number of sampled loci, (2) the more realistic demographic model that we used in our analysis, and (3) the absence of modern records of regular North America to Hawaii establishment events.

In conclusion, we have shown that monarchs colonized the Pacific as part of a single, recent out-of-North America expansion event, with at least two subsequent expansions out of Hawaii. Furthermore, we show that the loss of migration coupled with strong genetic drift can generate strong patterns of differentiation between monarch populations at the scale of islands within an archipelago, as seen in comparisons between the islands of Guam and Rota. This is in stark contrast to the continent-wide genetic panmixia within North America, and suggests that future studies must sample at both fine and coarse spatial scales if they hope to recover accurate population structures. Finally, understanding the magnitude of genomic, phenotypic, and ecological differentiation between migratory North American monarchs and populations in outlying U.S. states/territories (including Hawaii, American Samoa, and the Mariana Islands) has important conservation implications as the U.S. Fish and Wildlife Service considers the concept of adaptive capacity in their ongoing decision-making process of whether to list the monarch under the U.S. Endangered Species Act (Freedman et al. 2020b, USFWS).

**Data Accessibility**

All scripts used for analysis are available at <https://github.com/hemstrow/F-H_2018>. Sequence data will be made available through NCBI upon publication.

**Author Contributions**

WBH, MGF, and MRM designed the research. MGF and MPZ provided samples used for sequencing. WBH and MGF performed data analysis. All authors contributed to writing and editing the manuscript.

**Funding**

WBH was funded through the University of California, Davis Animal Science Department. MGF received funding through the NSF Graduate Research Fellowship, the National Geographic Society, the Society for the Study of Evolution, and the NSF EAPSI program. Funding for sequencing was provided by the University of California, Davis Animal Science Department to MRM. MPZ was supported by the University of Queensland, and SRR was supported by the David and Lucile Packard Foundation.

**Acknowledgments**

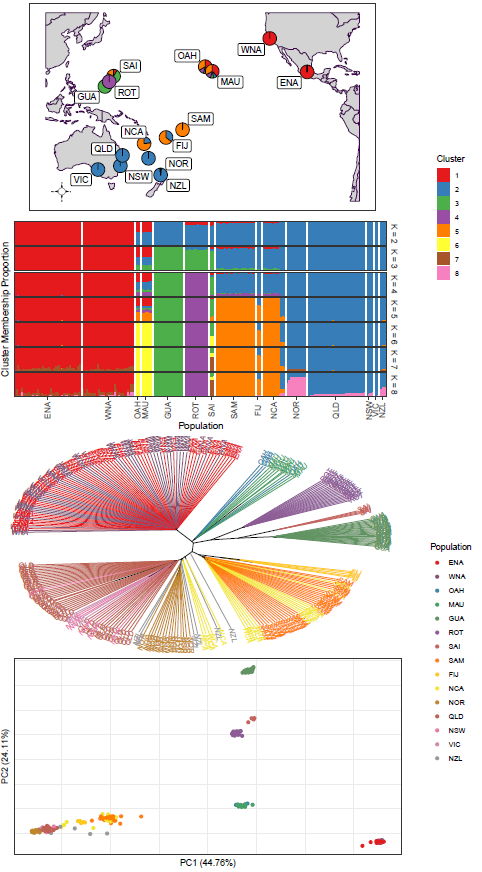
We are very grateful to the following people for providing samples used for sequencing: Cheryl Dean and Jessica Aguilar (Hawaii), Louie Yang (western North America), Tyler Flockhart, Ryan Norris, and Samantha Knight (eastern North America). Hugh Dingle, Haldre Rogers, Dan Fagin, and Ali Kerr assisted with monarch collection in Guam. We especially thank Sean O’Rourke for assistance with genomic library preparation.

**References**

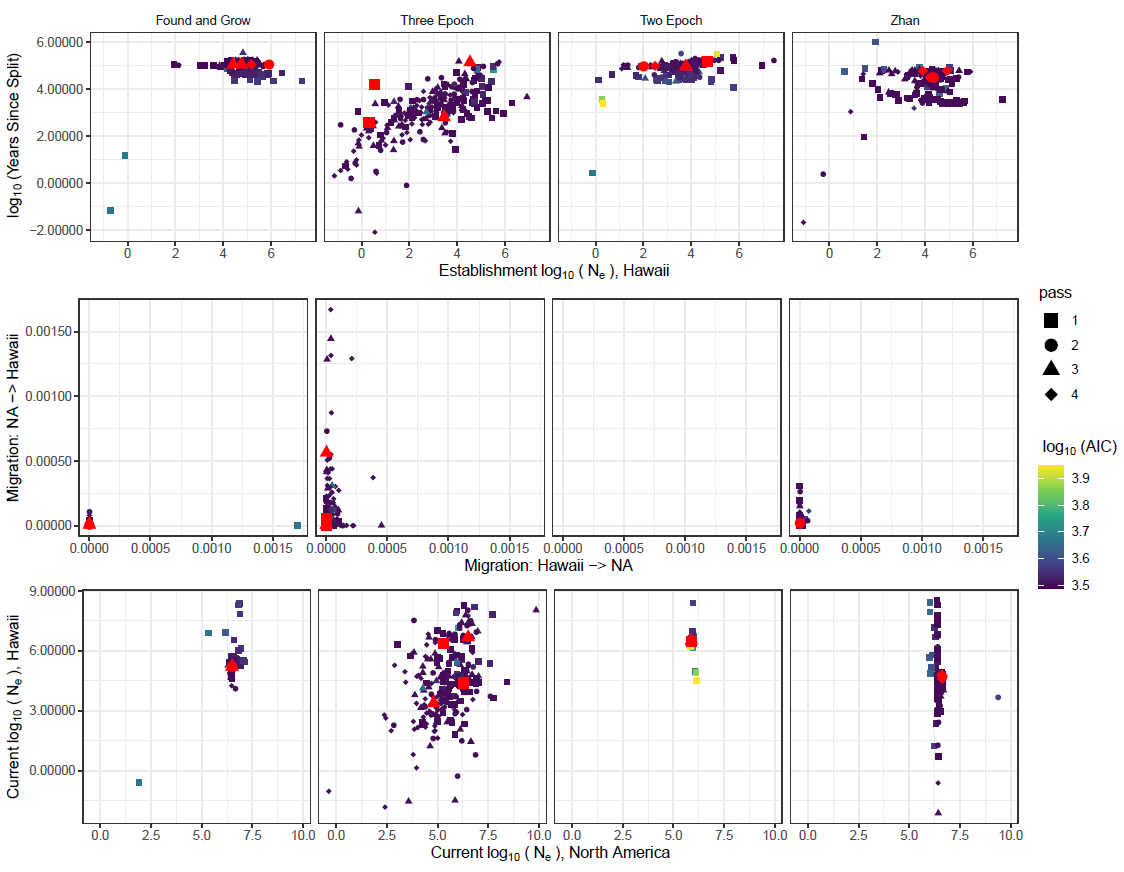
1. **Ackery, P. R., and R. I. Vane-Wright. 1984.** *Milkweed butterflies: Their Cladistics and Biology*. Ithaca (NY): Cornell University Press.
2. **Adriaensen, F., and A. A. Dhondt**. **1990**. Population dynamics and partial migration of the European robin (*Erithacus rubecula*) in different habitats. J. Anim. Ecol. 59: 1077–1090.
3. **Ali, O. A., S. M. O’Rourke, S. J. Amish, M. H. Meek, G. Luikart, C. Jeffres, and M. R. Miller**. **2016**. RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. Genetics. 202: 389–400.
4. **Alvial, I. E., H. A. Vargas, M. Marinov, C. Esquivel, J. Araya, R. Araya-Donoso, I. Vila, and D. Véliz**. **2018**. Isolation on a remote island: genetic and morphological differentiation of a cosmopolitan odonate. Heredity. 122: 893–905.
5. **Berg, J. E., M. Hebblewhite, C. C. St. Clair, and E. H. Merrill**. **2019**. Prevalence and mechanisms of partial migration in ungulates. Frontiers in Ecology and Evolution. 7: 325.
6. **Bettin, O., C. Cornejo, P. J. Edwards, and R. Holderegger**. **2007**. Phylogeography of the high alpine plant *Senecio halleri* (Asteraceae) in the European Alps: in situ glacial survival with postglacial stepwise dispersal into peripheral areas. Mol. Ecol. 16: 2517–2524.
7. **Billings, J.** **2019**. Opening a window on southwestern monarchs: Fall migrant monarch butterflies, *Danaus plexippus* (L.), tagged synchronously in southeastern Arizona migrate to overwintering regions in either southern California or central Mexico. J. Lepid. Soc. 73: 257–267.
8. **Brower, A. V. Z., and T. M. Boyce**. **1991**. Mitochondrial DNA variation in monarch butterflies. Evolution. 45: 1281–1286.
9. **Chapman, B. B., C. Brönmark, J.-Å. Nilsson, and L.-A. Hansson**. **2011**. The ecology and evolution of partial migration. Oikos. 120: 1764–1775.
10. **Charles, K. L., R. C. Bell, D. C. Blackburn, M. Burger, M. K. Fujita, V. Gvoždík, G. F. M. Jongsma, M. T. Kouete, A. D. Leaché, and D. M. Portik**. **2018**. Sky, sea, and forest islands: Diversification in the African leaf-folding frog *Afrixalus paradorsalis* (Anura: Hyperoliidae) of the Lower Guineo-Congolian rain forest. Journal of Biogeography.
11. **Clarke, A. R., and M. P. Zalucki**. **2004**. Monarchs in Australia: on the winds of a storm? Biol. Invasions. 6: 123–127.
12. **Cohen, E. B., J. A. Hostetler, M. T. Hallworth, C. S. Rushing, T. S. Sillett, and P. P. Marra**. **2018**. Quantifying the strength of migratory connectivity. Methods Ecol. Evol. 9: 513–524.
13. **Dapporto, L., A. Cini, M. Menchetti, R. Vodă, S. Bonelli, L. P. Casacci, V. Dincă, S. Scalercio, J. C. Hinojosa, H. Biermann, and Others**. **2017**. Rise and fall of island butterfly diversity: Understanding genetic differentiation and extinction in a highly diverse archipelago. Diversity and Distributions. 23: 1169–1181.
14. **Dingle, H., M. P. Zalucki, W. A. Rochester, and T. Armijo-Prewitt**. **2005**. Distribution of the monarch butterfly, *Danaus plexippus* (L.) (Lepidoptera: Nymphalidae), in western North America. Biol. J. Linn. Soc. Lond. 85: 491–500.
15. **Edwards, A. W.** **1971**. Distances between populations on the basis of gene frequencies. Biometrics. 27: 873–881.
16. **Excoffier, L., M. Foll, and R. J. Petit**. **2009**. Genetic consequences of range expansions. Annu. Rev. Ecol. Evol. Syst. 40: 481–501.
17. **Fernández-Haeger, J., D. Jordano, and M. P. Zalucki. 2015.** Monarchs across the Atlantic Ocean. In *Monarchs in a Changing World: Biology and Conservation of an Iconic Butterfly* (eds KS Oberhauser, KR Nail, S Altizer), pp. 247-256. Ithaca (NY): Cornell University Press.
18. **Francis, R. M.** **2017**. pophelper: an R package and web app to analyse and visualize population structure. Mol. Ecol. Resour. 17: 27–32.
19. **Freedman, M. G., H. Dingle, C. A. Tabuloc, J. C. Chiu, L. H. Yang, and M. P. Zalucki**. **2018**. Non-migratory monarch butterflies, *Danaus plexippus* (L.), retain developmental plasticity and a navigational mechanism associated with migration. Biol. J. Linn. Soc. Lond. 123: 265–278.
20. **Freedman, M. G., H. Dingle, S. Y. Strauss, and S. R. Ramírez**. **2020a**. Two centuries of monarch butterfly collections reveal contrasting effects of range expansion and migration loss on wing traits. Proc. Natl. Acad. Sci. U. S. A. 117: 28887–28893.
21. **Freedman, M. G. , J. De Roode, M. Forister, M. Kronforst, A. Pierce, C. Schultz, O. Taylor, and E. Crone**. **2020b**. Are eastern and western monarch butterflies distinct populations? A review of evidence for ecological, phenotypic, and genetic differentiation and implications for conservation. Preprints: 10.20944/preprints202009.0353.v1
22. **Gao, B., J. Hedlund, D. R. Reynolds, B. Zhai, G. Hu, and J. W. Chapman**. **2020**. The “migratory connectivity” concept, and its applicability to insect migrants. Mov. Ecol. 8: 48.
23. **Gómez-Bahamón, V., R. Márquez, A. E. Jahn, C. Y. Miyaki, D. T. Tuero, O. Laverde-R, S. Restrepo, and C. D. Cadena**. **2020**. Speciation associated with shifts in migratory behavior in an avian radiation. Curr. Biol. 30: 1312–1321.e6.
24. **Gong, X., E. R. Davenport, D. Wang, and A. G. Clark**. **2019**. Lack of spatial and temporal genetic structure of Japanese eel (*Anguilla japonica*) populations. Conserv. Genet. 20: 467–475.
25. **Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante**. **2009**. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5: e1000695.
26. **Haag-Liautard, C., M. Dorris, X. Maside, S. Macaskill, D. L. Halligan, D. Houle, B. Charlesworth, and P. D. Keightley**. **2007**. Direct estimation of per nucleotide and genomic deleterious mutation rates in Drosophila. Nature. 445: 82–85.
27. **Hewitt, G. M.** **1996**. Some genetic consequences of ice ages, and their role in divergence and speciation. Biol. J. Linn. Soc. Lond. 58: 247–276.
28. **Hughes, J. M., and M. P. Zalucki**. **1984**. Genetic variation in a continuously breeding population of *Danaus plexippus* L. (Lepidoptera: Nymphalidae). Heredity. 52: 1–7.
29. **Ibrahim, K. M., R. A. Nichols, and G. M. Hewitt**. **1996**. Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. Heredity. 77: 282–291.
30. **Irwin, D. E., J. H. Irwin, and T. B. Smith**. **2011**. Genetic variation and seasonal migratory connectivity in Wilson’s warblers (*Wilsonia pusilla*): species-level differences in nuclear DNA between western and eastern populations. Mol. Ecol. 20: 3102–3115.
31. **Jakobsson, M., and N. A. Rosenberg**. **2007**. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 23: 1801–1806.
32. **James, D. G. 1993.** Migration biology of the monarch butterfly in Australia. In *Biology and Conservation of the Monarch Butterfly* (eds SB Malcolm, MP Zalucki), pp. 189-200. Los Angeles (CA): Los Angeles County Museum of Natural History.
33. **James, D. G., and T. A. James**. **2019**. Migration and overwintering in Australian monarch butterflies (*Danaus plexippus* (L.) (Lepidoptera: Nymphalidae): A review with new observations and research needs. The Journal of the Lepidopterists’ Society.
34. **Keightley, P. D., A. Pinharanda, R. W. Ness, F. Simpson, K. K. Dasmahapatra, J. Mallet, J. W. Davey, and C. D. Jiggins**. **2015**. Estimation of the spontaneous mutation rate in Heliconius melpomene. Mol. Biol. Evol. 32: 239–243.
35. **Kondo, B., J. L. Peters, B. B. Rosensteel, and K. E. Omland**. **2008**. Coalescent analyses of multiple loci support a new route to speciation in birds. Evolution. 62: 1182–1191.
36. **Korneliussen, T. S., A. Albrechtsen, and R. Nielsen**. **2014**. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics. 15: 356.
37. **Kuchta, S. R., and A.-M. Tan**. **2005**. Isolation by distance and post-glacial range expansion in the rough-skinned newt, *Taricha granulosa*. Mol. Ecol. 14: 225–244.
38. **Li, H., and R. Durbin**. **2009**. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25: 1754–1760.
39. **Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup**. **2009**. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25: 2078–2079.
40. **Linderoth, T. P.** **2018**. Identifying population histories, adaptive genes, and genetic duplication from population-scale next generation sequencing. Ph.D. thesis: <https://escholarship.org/uc/item/5kp4q40k>[.](http://paperpile.com/b/tNxuHC/OQZa)
41. **Lyons, J. I., A. A. Pierce, S. M. Barribeau, E. D. Sternberg, A. J. Mongue, and J. C. De Roode**. **2012**. Lack of genetic differentiation between monarch butterflies with divergent migration destinations. Mol. Ecol. 21: 3433–3444.
42. **Mantel, N.** **1967**. The detection of disease clustering and a generalized regression approach. Cancer Res. 27: 209–220.
43. **Menz, M. H. M., D. R. Reynolds, B. Gao, G. Hu, J. W. Chapman, and K. R. Wotton**. **2019**. Mechanisms and consequences of partial migration in insects. Frontiers in Ecology and Evolution. 7: 403.
44. **Morris, G. M., C. Kline, and S. M. Morris**. **2015**. Status of *Danaus plexippus* population in Arizona. J. Lepid. Soc. 69: 91–107.
45. **Nei, M., T. Maruyama, and R. Chakraborty**. **1975**. The bottleneck effect and genetic variability in populations. Evolution. 29: 1–10.
46. **Paradis, E., and K. Schliep**. **2019**. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 35: 526–528.
47. **Peter, B. M., and M. Slatkin**. **2013**. Detecting range expansions from genetic data. Evolution. 67: 3274–3289.
48. **Pfeiler, E., N. O. Nazario-Yepiz, F. Pérez-Gálvez, C. A. Chávez-Mora, M. R. L. Laclette, E. Rendón-Salinas, and T. A. Markow**. **2017**. Population genetics of overwintering monarch butterflies, *Danaus plexippus* (Linnaeus), from central Mexico inferred from mitochondrial DNA and microsatellite markers. J. Hered. 108: 163–175.
49. **Pierce, A. A., M. P. Zalucki, M. Bangura, M. Udawatta, M. R. Kronforst, S. Altizer, J. F. Haeger, and J. C. de Roode**. **2014a**. Serial founder effects and genetic differentiation during worldwide range expansion of monarch butterflies. Proc. Biol. Sci. 281.
50. **Pierce, A. A., J. C. de Roode, S. Altizer, and R. A. Bartel**. **2014b**. Extreme heterogeneity in parasitism despite low population genetic structure among monarch butterflies inhabiting the Hawaiian islands. PLoS ONE.
51. **Pierce A. A., S. Altizer , N. L. Chamberlain, M. R. Kronforst, and J. C. de Roode. 2015.** Unraveling the mysteries of monarch migration and global dispersal through molecular genetic techniques. [In *Monarchs in a Changing World: Biology and Conservation of an Iconic Butterfly* (eds KS Oberhauser, KR Nail, S Altizer), pp. 257-267.](http://paperpile.com/b/tNxuHC/UJ0f) Ithaca (NY): Cornell University Press.
52. **Portik, D. M., A. D. Leaché, D. Rivera, M. F. Barej, M. Burger, M. Hirschfeld, M.-O. Rödel, D. C. Blackburn, and M. K. Fujita**. **2017**. Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. Mol. Ecol. 26: 5245–5263.
53. **Prince, D. J., S. M. O’Rourke, T. Q. Thompson, O. A. Ali, H. S. Lyman, I. K. Saglam, T. J. Hotaling, A. P. Spidle, and M. R. Miller**. **2017**. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. Sci Adv. 3: e1603198.
54. **Rousset, F.** **2008**. genepop’007: a complete re-implementation of the genepop software for Windows and Linux. Mol. Ecol. Resour. 8: 103–106.
55. **Saitou, N., and M. Nei**. **1987**. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406–425.
56. **Shephard, J. M., J. M. Hughes, and M. P. Zalucki**. **2002**. Genetic differentiation between Australian and North American populations of the monarch butterfly *Danaus plexippus* (L.) (Lepidoptera: Nymphalidae): an exploration using allozyme electrophoresis. Biol. J. Linn. Soc. Lond. 75: 437–452.
57. **Skotte, L., T. S. Korneliussen, and A. Albrechtsen**. **2013**. Estimating individual admixture proportions from next generation sequencing data. Genetics. 195: 693–702.
58. **Slatkin, M., and L. Excoffier**. **2012**. Serial founder effects during range expansion: a spatial analog of genetic drift. Genetics. 191: 171–181.
59. **Stefanescu, C., M. Alarcón, and A. Avila**. **2007**. Migration of the painted lady butterfly, *Vanessa cardui*, to north-eastern Spain is aided by African wind currents. J. Anim. Ecol. 76: 888–898.
60. **Tajima, F.** **1989**. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123: 585–595.
61. **Talla, V., A. A. Pierce, K. L. Adams, T. J. B. de Man, S. Nallu, F. X. Villablanca, M. R. Kronforst, and J. C. de Roode**. **2020**. Genomic evidence for gene flow between monarchs with divergent migratory phenotypes and flight performance. Mol. Ecol. 29: 2567–2582.
62. **Tenger-Trolander, A., W. Lu, M. Noyes, and M. R. Kronforst**. **2019**. Contemporary loss of migration in monarch butterflies. Proc. Natl. Acad. Sci. U. S. A. 116: 14671–14676.
63. **Tiedemann, R., K. B. Paulus, M. Scheer, K. G. Von Kistowski, K. Skírnisson, D. Bloch, and M. Dam**. **2004**. Mitochondrial DNA and microsatellite variation in the eider duck (*Somateria mollissima*) indicate stepwise postglacial colonization of Europe and limited current long-distance dispersal. Mol. Ecol. 13: 1481–1494.
64. **USFWS (United States Fish and Wildlife Service)**. Assessing the status of the monarch butterfly. https://www.fws.gov/savethemonarch/ssa.html.
65. **Vane-Wright, R. I. 1993.** The Columbus hypothesis: an explanation for the dramatic 19th century range expansion of the monarch butterfly. In *Biology and Conservation of the Monarch Butterfly* (eds SB Malcolm, MP Zalucki), pp. 179-188. Los Angeles (CA): Los Angeles County Museum of Natural History.
66. **Waples, R. S., D. J. Teel, J. M. Myers, and A. R. Marshall**. **2004**. Life-history divergence in Chinook salmon: historic contingency and parallel evolution. Evolution. 58: 386–403.
67. **Wise, K. A. J.** **1980**. Monarch butterfly dispersal in New Zealand. Records of the Auckland Institute and Museum. 17: 157–173.
68. **Zalucki, M. P., J. M. Hughes, and P. A. Carter**. **1987**. Genetic variation in *Danaus plexippus* L.: Habitat selection or differences in activity times? Heredity. 59: 213–221.
69. **Zalucki, M. P., and A. R. Clarke**. **2004**. Monarchs across the Pacific: the Columbus hypothesis revisited. Biol. J. Linn. Soc. Lond. 82: 111–121.
70. **Zhan, S., and S. M. Reppert**. **2013**. MonarchBase: the monarch butterfly genome database. Nucleic Acids Res. 41: D758–63.
71. **Zhan, S., W. Zhang, K. Niitepõld, J. Hsu, J. F. Haeger, M. P. Zalucki, S. Altizer, J. C. de Roode, S. M. Reppert, and M. R. Kronforst**. **2014**. The genetics of monarch butterfly migration and warning colouration. Nature. 514: 317–321.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Population** | **# Samples** | **Tajima's D** | **H­O** | **π** | **Het/Hom** |
| **NAM** | 83 | -1.92 | 0.055 | 0.064 | 0.059 |
| **HAW** | 9 | -0.211 | 0.048 | 0.055 | 0.051 |
| **GUA** | 19 | 0.091 | 0.031 | 0.032 | 0.032 |
| **ROT** | 16 | 0.388 | 0.035 | 0.038 | 0.037 |
| **SAI** | 4 | 0.326 | 0.022 | 0.025 | 0.023 |
| **QLD** | 15 | 0.349 | 0.042 | 0.044 | 0.043 |
| **NSW** | 5 | 0.433 | 0.038 | 0.042 | 0.039 |
| **VIC** | 2 | 0.899 | 0.041 | 0.043 | 0.042 |

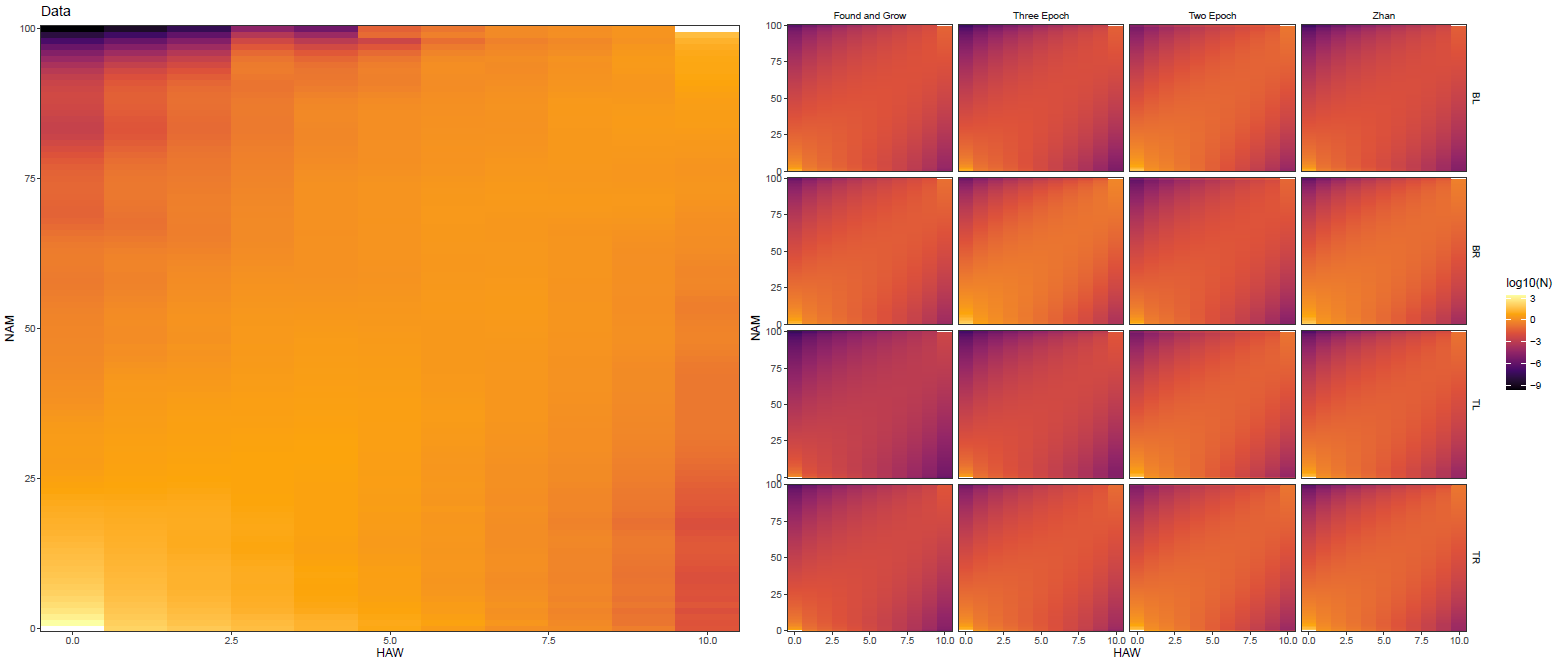
**Table 1 -** Number of samples remaining after filtering, Tajima’s D, Observed Heterozygosity (HO), nucleotide diversity (π), and the average ratio of Heterozygous to Homozygous sites across all individuals in each population.



**Figure 1 –** Relatedness among sampled populations. Top to bottom: (a) Map of sampled populations, with pie charts reflecting results from NGSadmix. (b) NGSadmix plots showing the proportion of ancestry across clustering values between k = 2 and k = 9. At k = 5, Hawaii reflects a mixture of ancestry comprising North American, Mariana Islands, and southwestern Pacific samples. At k = 6, Hawaii becomes its own cluster. At values beyond k = 6, populations are subdivided. (c) Neighbor joining tree. (d) Principal component analysis largely recapitulates the geographical distribution of samples, with PC1 explaining 44.7% of variation and corresponding to the east-west axis of differentiation. Abbreviations in labels are as follows: ENA = Eastern North America, WNA = Western North America, OAH = Oahu (Hawaii), MAU = Maui (Hawaii), GUA = Guam, ROT = Rota, SAI = Saipan, SAM = Samoa, FIJ = Fiji, NCA = New Caledonia, NOR = Norfolk Island, QLD = Queensland, NSW = New South Wales, VIC = Victoria, NZL = New Zealand.



**Figure 2 -** Results of the dadi optimization runs for the (left to right) *Found and Grow*, *Three Epoch*, *Two epoch*, and *Zhan* models demographic models. (Top) Hawaii establishment effective size and years since establishment. (Middle) Migration rates from North America to Hawaii and vice versa. (Bottom) Current effective size estimates in North America and Hawaii. Red dots mark the runs with the lowest AIC scores in each quadrant of the respective parameter space based on the (a), corresponding to the heatmaps in Figure 3 and the residuals in Figure S5.



**Figure 3 –** Observed (Left) and estimated (Right) derived site frequency spectra for the *Found and Grow*, *Three Epoch*, *Two Epoch*, and *Zhan* models. Estimated spectra based on the parameters from the runs with the lowest AIC score from each quadrant of the establishment time/founding population size parameter space for each model are shown for comparison (BL: bottom left, BR: bottom right, TL: top left, TR: top right), corresponding to the points marked in red in Figure 2.