**Population genetics of a recent range expansion in monarch butterflies**

Will Hemstrom1\*, Micah Freedman2,3\*, Myron Zalucki4, Sharon Strauss2,3, Santiago Ramírez2,3, Tyler Flockhart5, Ryan Norris6, Michael Miller1

1. Department of Animal Science, UC Davis
2. Department of Evolution and Ecology, UC Davis
3. UC Davis Center for Population Biology
4. School of Biological Sciences, University of Queensland, Australia
5. University of Maryland
6. University of Guelph

\*Authors contributed equally

**Introduction**

Understanding the patterns and processes that define species’ geographic ranges is a fundamental goal in ecology and evolutionary biology. Geographic ranges are generally a reflection of species’ fundamental niches, with changes in range size driven by gradual changes in climate (Hewitt 1999, Davis and Shaw 2001) and rare long-distance dispersal events (Gillespie et al. 2012). Over the past 200 years, the tempo and magnitude of species range expansions has greatly increased (Helmus et al. 2014, others). This is primarily the result of deliberate or accidental introductions of species associated with human agriculture and commerce, although many recent examples also highlight the role of anthropogenic climate change in mediating range shifts (Parmesan and Yohe 2003, others). Studying the population genetics of range-expanding species is important because it informs our understanding of important features such as the timing of expansion, the number of expansion / introduction events, and the amount of ongoing gene flow between ancestral and derived populations (Peter and Slatkin 2015).

Species that undergo range expansions often have distinctive patterns of population genetic structure, with decreasing relatedness and increasing contributions of genetic drift in populations further from the original source population (Hewitt 1996, Excoffier et al. 2009). One commonly encountered form of range expansion is stepwise serial stepwise dispersal, 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 as diverse as eider ducks (*Somateria mollissima*) (Tiedemann et al. 2004), ragwort (*Senecio halleri*) (Bettin et al. 2007), and rough-skinned newts (*Taricha granulosa*) (Kuchta and Tan 2005). The out-of-Africa expansion of *Homo sapiens* is also characterized by serial stepwise dispersal (Ramachandran et al. 2005, Liu *et al.* 2006).

One such species that has undergone a dramatic range expansion over its recent evolutionary history is the monarch butterfly (*Danaus plexippus* (L.)). Evidence suggests that the monarch historically occupied Central America and the southern United States before undergoing a large demographic expansion around 20,000 years ago (Zhan et al. 2014, Pfeiller et al. 2016). This demographic expansion likely coincided with the end of the last ice age and glacial retreat in North America, which enabled colonization of temperate areas by the monarch’s *Asclepias* host plants and likely set the stage for the onset of continent-scale long-distance migration in the monarch. Much more recently, the monarch has become established around the globe in a number of independent out-of-North America expansion events (Zhan et al. 2014). This includes a southern expansion that involved establishment in South America and the Caribbean (likely thousands or tens of thousands of year ago), an eastward expansion across the Atlantic and into the Iberian Peninsula, and a westward expansion across the Pacific. In this paper, we focus exclusively on the monarch’s range expansion across the Pacific.

Historical records suggest that the monarch crossed the Pacific quite recently, with the earliest positive records of monarch occurrences coming from the 1840s in Hawaii (Vane-Wright 1993, Zalucki and Clarke 2004). By 1871, the monarch had reached Australia and was established on nearly every major Pacific island group by 1900. Some authors have attributed the recency of the monarch’s appearance in these locations to the “Columbus hypothesis,” which posits that the clearing of forests in the eastern and midwestern United States during the 1800s prompted a massive increase in the North American monarch’s population size and scope of migration (Vane-Wright 1993). However, demographic reconstructions using whole genome sequence data indicate that the monarch’s out-of-North America expansion events happened much more distantly, perhaps as long as 2,000-3,000 years ago (Zhan et al. 2014). Thus, there is still disagreement between demographic models and historical records about the timing of the monarch’s Pacific expansion.

Currently available population genetic data suggest that the monarch’s recent global range expansion happened in a serial stepwise fashion (Pierce et al. 2014a, Zhan et al. 2014). Serial expansion is indicative of natural expansion wave, rather than a series of independent and deliberate human introductions, as suggested by Zalucki and Clarke (2004). Still, there are a number of unsampled populations in the Pacific that might improve our understanding of establishment timing and direction.

Finally, in contrast to their migratory North American ancestors, nearly all Pacific island populations have become fully sedentary, year-round breeding populations. Little is known about how this contemporary loss of migration has affected fine-scale patterns of population differentiation in monarchs. One study has addressed this question: Pierce et al. (2014b) used microsatellites and showed that monarchs from Hawaii show little differentiation among islands. However, the conclusions of this study were based on only 11 variable loci from one island group. Thus, the degree to which loss of migration shapes fine-scale patterns of population differentiation remains unresolved.

In this study, we use reduced-representation whole genome sequencing across a sample of approximately 280 monarch butterflies to understand (1) patterns of relatedness among Pacific and North American populations, (2) expansion timing and amount of ongoing gene flow from North America, and (3) genetic differentiation within expansion populations. Our dataset contains tens of thousands of variable sites from monarchs in 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. We find support for a stepwise pattern of dispersal across the Pacific, but with a previously uncharacterized westward expansion from Hawaii into the Mariana Islands. Estimates for the timing of the monarch’s establishment in the Pacific are concordant with a recent expansion, but with high uncertainty around the precise timing of this event. Gene flow from North America to Hawaii appears to be unidirectional, with low levels of ongoing North America to Hawaii gene flow. 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 panmictic sample across the entire continent. Together, our data argue in favor of deferring to historical records to infer the history of the monarch’s range expansion and also provide a fascinating example of how migratory status can influence patterns of genetic isolation.

**Methods**

*Sample collection and storage*

Monarchs were collected as either larvae or adult butterflies from locations around the world between 1990 – 2017. When possible, monarchs were collected over as broad of a spatial and temporal window as possible to minimize the chances of sampling full or half sibs. Monarchs were stored in ethanol and keep either at room temperature or -20C freezers prior to DNA extraction. For a summary of monarchs included in our sequencing, see Figure 1 and Table S1.

*Sample preparation and sequencing*

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 Pst1 restriction enzyme according to Ali et al (2016) and sequenced using 100bp paired-end sequencing on an Illumina Hi-Seq 2500.

*Sequence alignment, filtering, and genotype calling*

Raw sequence data was aligned to version 3 of the monarch butterfly genome assembly (Zhan et al., 2011) using the Burrows-Wheeler Alignment algorithm (Li & Durbin, 2009). Sequence data was sorted and filtered for PCR duplicates and improper pairs using SAMtools (Li et al., 2009). Genotypes were then called using the SAMtools genotype likelihood model (Li 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 posterior genotype probability cutoff of 0.95, and a minimum minor allele frequency of 0.05 (Korneliussen et al., 2014). Genotypes were then filtered to remove any SNPs sequenced in less than 50% of samples or that had a heterozygote frequency of > 0.55 in order to remove SNPs likely to originate due to genome duplications using the snpR package (Hemstrom et al. *in prep*).

*Statistical Analysis*

The average number of pairwise differences (µ), observed heterozygosity (Ho), and fixation index (FST) between each pair of populations for each SNP was then calculated using the snpR package. FST was calculated according to Weir & Cockerham (1984).

In order to characterize basic population structure, we then created a neighbor-joining tree (Saitou & Nei, 1987) using the ape R package (Paradis et al., 2004). 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 using the same parameters as above (Korneliussen et al., 2014). For comparison, a sparse Non-Negative Matrix Factorization algorithm was also used to generate individual ancestry coefficients for each individual for between 1 and 12 putative population clusters (k) (Frichot et al., 2014) using the LEA R package (Frichot & François, 2015). Each value of k was run 10 times, and the results were collapsed into consensus plots using CLUMPP (Jakobsson & Rosenberg, 2007). The pophelper (Francis, 2017) and snpR (Hemstrom et al. *in prep*) R packages were used to simplify this analysis. Lastly, NGSadmix was also used to calculate ancestry proportions for each individual at each value of k from 1 to 12 (Skotte et al., 2013) and a Principal Component Analysis (PCA) was conducted on the called genotypic data for comparison.

To describe the patterns of establishment and migration between North America and the Pacific, the demographic reconstruction program δaδi (dadi) was used (Gutenkunst et al., 2009) 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 SNP was within 10,000bp of any other SNP. The resulting 11,384 SNPs were polarized via reference to whole genome sequence data of the best sequenced monarch sister taxa *Danaus erippus* individual from Zhan et al. (2014) by alignment to the monarch genome as described above. A range of possible models were fit to the observed data, including the model described by Zhan et al. (2014) for the same comparison, a similar model that allowed for an additional period of growth prior to the establishment of the Hawaiian population and another following establishment, the two-dimensional models described by Portik et al. (2017), and modifications on the later models with logistic rather than exponential growth functions. These models are depicted in more detail in Figure X. To optimize the fitted models, the sequential step-down parameter permutation approach described by Portik et al. (2017) was used. The number of model runs and iterations per step are listed in Table X. Unlike Portik et al. (2017), however, the starting parameters for each sequential run were set 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.

To quantify the direction and strength of population spread across the Pacific, the directionality index () was calculated for each pairwise combination of the North American, Hawaiian, Queensland, and Guamanian populations according to Peter & Slatkin (2013) using the snpR package (Hemstrom et al *in prep*). The polarized site-frequency spectra used in these calculations were created using the dadi dataset described above using the methods described by Gutenkunst et al. (2009) as implemented in snpR (Hemstrom et al *in prep*).

**Results**

*Question 1: Overall patterns of relatedness*

PC1 explained (\_\_%) of the overall variance and separated North American from Pacific Island samples. PC2 explained (\_\_%) of variance and split Pacific Island populations into two out-of-Hawaii expansions (Figure 2a). North American monarchs formed a single panmictic populations in all analyses. Consistent with patterns of natural range expansion, we find decreasing relatedness to the ancestral North American population with increasing distance.

sNMF and NGSrelate showed a generally similar pattern. At k = 2, North American and Pacific Island populations were pulled apart. At k = 3, Guam was separately from all other Pacific population. At k = 4, samples from Rota were assigned their own cluster. At k = 5, Samoa, Fiji, and New Caledonia were assigned their own cluster, and at k = 6, Hawaii was assigned a cluster. Values of k = 7 and higher only subdivided populations, and even at values as high as k = 10, eastern and western North American populations did not form discrete clusters (Figure 2b).

*Question 2: Timing of establishment and patterns of ongoing gene flow*

Models suggest a recent colonization of Hawaii from North America. However, the exact timing is difficult to pinpoint and is sensitive to demographic scenarios specified during model specification. Models consistently predict a large effective population size in North America of approximately (); interestingly, the inferred founding population size in Hawaii was inferred to be quite high, approximately ().

There appears to be a very small amount of ongoing gene flow from North America to Hawaii, on the order of \_\_\_\_ individuals per generation. However, there is no support for Hawaii > North America gene flow.

*Question 3: Patterns of differentiation within expansion populations*

Samples from the Mariana Islands (especially the well-sampled Guam and Rota populations) appear to form distinct populations, despite their extremely close physical proximity. By contrast, populations within Hawaii (Maui and Oahu) and Australia (Queensland, New South Wales, and Victoria) do not show strong patterns of differentiation. Likewise, we find support for a single panmictic North American population that encompasses both eastern and western populations.

**Discussion**

Our analysis suggests a population genetic scenario consistent with a recent natural range expansion characterized by serial stepwise dispersal across the Pacific in monarchs. With increasing distance, Pacific populations become more distantly related to the ancestral North American population. We find evidence for two independent expansions upon establishment in Hawaii, with a previously uncharacterized westward expansion from Hawaii into the Mariana Islands. The other previously unsampled population in our data, from Norfolk Island, appears to be part of the monarch’s southwestward Pacific expansion and generally groups with samples from Australia and New Zealand. Our results are broadly concordant with analyses by Zhan et al. (2014) and Pierce et al. (2014) but provide a higher resolution picture of the monarch’s pattern of establishment in the Pacific.

We found that monarchs in the Mariana Islands represent a distinct expansion event within the Pacific. Historical records provided in Zalucki and Clarke (2004) indicate that the monarch’s establishment date in the Marianas was likely around 1900, although there are museum specimens from Guam dating back to at least 1887 and from Saipan that date to at least 1883 (M. Freedman, pers. obs.). It is likely, based on historical records and geography, that monarchs in the Marianas are themselves descended from populations in the Marshall Islands and/or Micronesia, where there are monarch records dating back to at least 1860 (Zalucki and Clarke 2004). These islands have yet to be sampled in population genetic analyses, although there are extant monarch populations present there (British Museum of Natural History collection; J. Tennant, pers. comm.).

Within the Mariana Islands, we found a strong pattern of differentiation between islands, especially between the nearby islands of Guam and Rota. This pattern is striking because of their extremely close 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 2,000 km apart, formed a single panmictic population. This result 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, reviewed in Pierce et al. 2015), although it provides the strongest test of North American population structure to date\*.

This pattern of 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. Panmixia over large spatial scales in common in other long-distance migratory species (references from bats, birds, salmon), though monarchs provide a unique opportunity to compare patterns of population structure in migratory and non-migratory populations. This result also highlights that Pacific populations, despite comprising a single range expansion event, have spent sufficient time in isolation to become distinct evolutionary entities that can be treated as semi-independent replicates in comparative studies (e.g. Freedman et al., in review).

In contrast to populations within the Mariana Islands, samples from Hawaiian and Australian monarchs do not provide indications of fine-scale population genetic differentiation that might be expected in non-migratory populations. Within Hawaii, our samples from Maui and Oahu formed a single genetic cluster, consistent with the results of Pierce et al. (2014b). Likewise, within the Australian continent, samples from New South Wales and Victoria grouped with samples from Queensland. This result is consistent with early studies of broad- and fine-scale population genetic structure of Australian monarch butterflies using sampling from a small number of loci (Zalucki et al. 1987).

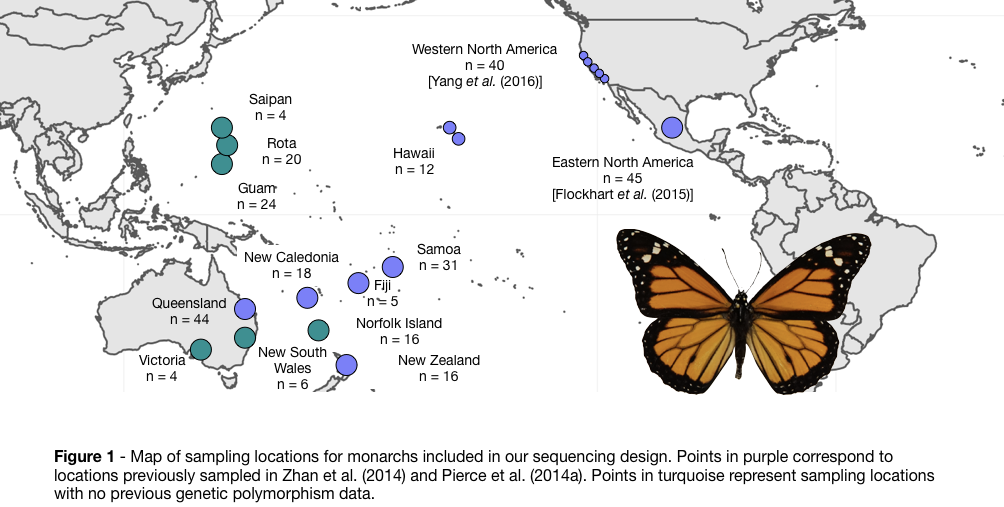
For Hawaiian monarchs, it is not immediately clear why the islands of Maui and Oahu do not form separate populations. One possibility is that prevailing winds promote gene flow between islands and that the orientation axis of the Hawaiian islands relative to prevailing wind direction differs from that seen in the Mariana Islands. Pacific monarchs are likely moved by wind patterns, and some authors have suggested that tropical cyclones may have promoted the monarch’s establishment in parts of the Pacific, including Australia (Clarke and Zalucki 2004). In the case of Australian monarchs, one possibility for the lack of differentiation across the continent is that Australian monarchs may in fact exhibit seasonal movement patterns akin to those seen in North American monarchs (James references). Although the scale of migration in Australia is likely relatively modest, Australian monarchs still exhibit seasonal movement patterns and migratory tendencies (James 1989, Freedman et al. 2018, Hemstrom et al. in prep) that might explain the lack of differentiation seen there.

Our demographic model results support the idea for a recent introduction of monarchs into Hawaii, with most models suggesting establishment timing around \_\_\_\_ generations ago. However, these models appear to be highly sensitive to specified input parameters, including population growth patterns upon establishment. Demographic models such as dadi may struggle with inferring very recent divergence patterns, as highlighted by \_\_\_\_\_. Given the sensitivity of these models, along with phenotypic evidence for contemporary differentiation between Pacific and North American populations, we suggest deferring to historical records in inferring approximate establishment timing for Pacific populations. Historical records also make more sense than demographic simulations when considering that monarchs require milkweed (*Asclepias*, *Gomphocarpus*, *Calotropis*) host plants that are almost certainly recent human introductions into the Pacific themselves. Finally, recent sequencing of insectary-reared monarch populations has shown that as little as 20 years of captive-breeding is sufficient to generate patterns of genetic divergence recapitulating those that we observed in Pacific populations (Tenger-Trolander et al. 2019).

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 between North America and Hawaii. Off-course migratory monarchs have been documented in relatively large numbers well outside of their North American range, with periodic reports of hundreds to thousands of butterflies in the United Kingdom in some years; these monarch records generally coincide with unusual weather patterns and are often correlated with aberrant records of migratory North America birds (reference). However, North America to the UK encompasses a much shorter distance than North America to Hawaii, and we are not aware of any comparable records of large-scale monarch influxes into Hawaii. A more likely scenario is that a very small number of founding individuals reached Hawaii and then rapidly expanded their numbers in an area with abundant host plants and naïve predators. A single gravid female monarch can lay more than 600 eggs in her lifetime (Oberhauser reference, pers. obs.), potentially mitigating some of the bottleneck effects associated with initial establishment (any references from other systems to support this?).

In contrast to variable estimates of establishment timing and founding population size, demographic models were very consistent in suggesting unidirectional migration from North America to Hawaii. Contemporary migration rates are still inferred to be low (on the order of 0.01 individuals per generation from NA to Hawaii). Migration from Hawaii to North America was suggested to be negligible, which makes sense in light of prevailing wind directions and loss of migratory behavior in Hawaiian monarchs. Our results thus contrast with those of Pierce et al (2014a), whose methods suggested much higher migration rates for both North America to Hawaii and vice versa. However, we are more confident in our results than those of Pierce et al. (2014) due to the much larger number of sampled loci and the more sophisticated demographic model that we used in our analysis.

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 panmixia that is maintained by seasonal migration within North America. Future studies of monarch population genetics would benefit from sampling additional likely establishment routes within the Pacific: these include populations in the far southern (French Polynesia, the Marquesas) and the far western (Taiwan, Hong Kong) Pacific. Likewise, denser sampling within other island groups across space and time might reveal whether the fine-scale differentiation we see within the Mariana Islands is the exception or the rule for non-migratory monarch populations. Finally, understanding the magnitude of genomic, phenotypic, and ecological differentiation between migratory North American monarchs and populations in outlying U.S. states/territories (Hawaii, American Samoa, the Mariana Islands) could have important conservation implications as the U.S. Fish and Wildlife Service considers a petition to list the monarch under the Endangered Species Act.



|  |  |  |  |
| --- | --- | --- | --- |
| **Population** | **Sampling Location** | **Sampling Year(s)** | **# Sequenced** |
| **North America** | Eastern North America (Mexican overwintering sites) | 2016 | 45 |
| Western North American (California overwintering sites) | 2015 | 40 |
| **Hawaii** | Maui | 2016 | 8 |
| Oahu | 2016 | 4 |
| **Mariana Islands** | Guam | 2015 | 24 |
| Rota | 2015 | 20 |
| Saipan | 2015 | 4 |
| **Fiji** | Viti Levu | 2009 | 5 |
| **Samoa** | Upolu | 2006, 2007, 2016 | 31 |
| **New Caledonia** | Grand Terre | 1991, 2006, 2010 | 18 |
| **Australia** | Queensland | 2016 | 44 |
| Victoria | 2016 | 4 |
| New South Wales | 2016 | 6 |
| **New Zealand** | North Island | 2007, 2011 | 6 |
| **Norfolk Island** | Norfolk Island | 2016 | 16 |
| **Total:** | | | **281** |
| Table S1 – Sampling locations and year for monarchs included in sequencing design | | | |