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

Will Hemstrom1\*, Micah Freedman2,3\*, Myron Zalucki4, 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

**Abstract**

**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 (Parmesan and Yohe 2003, Helmus et al. 2014). This is primarily the result of deliberate or accidental introductions of species associated with human agriculture and commerce (Hulme et al. 2009), although many recent examples also highlight the role of anthropogenic climate change in mediating range shifts (Parmesan and Yohe 2003, Dawe and Boutin 2016). 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, Henn et al. 2012).

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 (Ackery and Vane-Wright 1984, 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 monarchs 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, monarchs had reached Australia and were 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 microsatellite markers 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). For use in demographic reconstruction, 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, and a posterior genotype probability cutoff of 0.95 (Korneliussen et al., 2014). To reduce the likelihood that SNPs were in substantial physical linkage prior to demographic reconstruction, we then randomly selected SNPs such that no SNP was within 10kb of another using a custom R script. For all other analyses that relied on called genotypes, we instead 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 (Linderoth, 2018). Individuals sequenced at 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).

*Statistical Analysis*

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. We calculated FST using the R implementation of the GENEPOP software package (Rousset, 2008) with a minor allele frequency 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 Western and Eastern North American samples were lumped.

In order to characterize basic population structure, we then created a neighbor-joining tree (Saitou & Nei, 1987) using the ape R package v.5.0 (Paradis et al., 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 using the same parameters as above, save for a minor allele frequency cutoff of 0.05 (Korneliussen et al., 2014). A Principal Componenet Analysis (PCA) was also conducted using this dataset. For comparison, NGSadmix also 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 & 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 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. This 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 were fit to the observed data, including the models described in Charles et al. (2018) and Portik et al. (2017), variations on these models with logistic rather than exponential growth functions, the model described by Zhan et al. (2014) for the same comparison, and a similar model that allowed for an additional period of growth prior to the establishment of the Hawaiian population and another following establishment. The latter model is described in Figure S1. To optimize the fitted models, a variation on *dadi\_pipeline*, the sequential step-down parameter permutation approach described by Portik et al. (2017), was used. Unlike Portik et al. (2017), 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 S2. Individual optimization runs were killed if they took longer than 48hrs to complete, since these runs tended to proceed to take far longer to finish. Most runs completed in under 48hrs and are included in the results. 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). These values match those used by Zhan et al. (2014). Using a possibly more realistic generation time of roughly 7 generations per year results 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) results in large effective size estimates and more distant divergence times, for a net result of slightly more distant divergence times and larger effective sizes, but qualitatively similar results. In order to determine length of the considered genomic region, we multiplied the total number of bases sequenced after quality (but not SNP p-value) filtering by the ratio of SNPs in the final allele frequency spectrum to the total number of called, considered SNPs.

Among the large set of possible demographic models, the basic *found and grow* scenario, 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 new *three epoch found and grow* model, which involved multiple rounds of demographic expansion in the ancestral North American population, followed by colonization and growth in Hawaii had the lowest possible AIC score across all passes of the pipeline. As described above, the three epoch found and grow model is a more complex version of the model specified in Zhan *et al.* (2014). We report the results of both of these models and, where relevant, highlight discrepancies in the inferences that they produce.

To quantify the direction and strength of population spread across the Pacific, we calculated the directionality index () for each pairwise combination of the North American, Hawaiian, Queensland, Guam, Rota, Norfolk Island populations according to Peter & Slatkin (2013) using the snpR package (Hemstrom et al *in prep*). We created the polarized site-frequency spectra used in these calculations using the dadi dataset described above by projecting populations down to ten gene copies each using the methods described by Gutenkunst et al. (2009) as implemented in snpR (Hemstrom et al *in prep*). 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 monarch sister taxa *Danaus erippus* individual from Zhan et al. (2014) by alignment to the monarch genome as described above.

**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.

*Question 1: Overall patterns of relatedness*

Site counts PC1 explained 44.8% of the overall variance and separated North American from Pacific Island samples. PC2 explained 24.1% 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, as indicated by the directionality index () scores shown in Figure 3. Genetic diversity (π, HO, and Het/Hom), was highest in the North American populations, followed by Hawaii, Australia, and then the remaining Pacific Island populations, consistent with predictions of serial stepwise dispersal and strong bottlenecks (Table 1, Figure S2). Likewise, Tajima’s D is positive in all sites save for North America and Hawaii, indicating recent population bottlenecks in the former locations and population expansions in the latter (Table 1). Pairwise FST values were much lower between the different Australian samples than between the other populations. For each other population, the lowest FST value was vs. Hawaii. For Hawaii, the lowest FST value was vs. North America (see Table 2).

NGSrelate showed a generally similar pattern. At k = 2, North American and Pacific Island populations were pulled apart. Guam and Rota were separated from all other Pacific populations at k = 3 and 4, respectively. At k = 5, Samoa, Fiji, and New Caledonia were assigned their own cluster. At k = 6 Hawaii is assigned to a unique cluster; at k < 6, Hawaii appears to be intermediate between most other clusters. Lastly, Saipan was separated at k = 7. Values of k = 8 and higher only subdivided populations. At no k values did eastern and western North American populations form discrete clusters (Figure 2b).

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

For some parameters, models produced different inferences, with the more complicated *three epoch* model generally suggesting a much wider range of possible parameter values. For example, the simpler *found and grow* models were consistent in indicating a founding time of approximately 105 generations ago, while the *three epoch* models suggested a much broader range of establishment times that ranged between approximately 102 – 105 generations ago (Figure 4a, 4b). Similarly, the found and grow models were more consistent in predicting a large founding population of > 105 individuals, while the three epoch models suggested a found population size of between 10 – 105 individuals (Figure 4a, 4b). These models also differed in their estimates of the Ne for the Hawaiian population, with the found and grow model suggesting a large Ne of around 106 and the three epoch models generally producing estimates of Hawaiian Ne between 102 and 105 (Figure 4c, 4d).

For other parameters, the *found and grow* and *three epoch* models generated similar estimates. Both models suggest very low levels of contemporary migration between North America and Hawaii, with the *found and grow* models converging near 0 for both directions and the *three epoch* models generally suggesting migration rates of < 2.5 x 10-5 individuals per generation (Figure 4e, 4f). This finding accords with our intuition that trans-oceanic dispersal events in monarchs should be exceedingly rare events.

*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 evidence to date that North American monarchs form a single genetically panmictic population.

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 is common in other long-distance migratory taxa including bats (Peel et al. 2013), birds (Kraus et al. 2013), and eels (Als et al. 2011), 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 1993, Freedman et al. 2018, Hemstrom et al. in prep) that might explain the lack of differentiation seen there.

Summary statistics are consistent in indicating directional migration from North America > Hawaii, and then Hawaii > Guam and Hawaii > Australia. This pattern is reflected in both the positive directionality index measures (0.07, 0.08, and 0.05, respectively) and other summary statistics. 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.

Interpretation of our demographic model results is somewhat complicated. 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 expansion—has empirical support from other studies (Zhan et al. 2014, Pfeiller et al. 2016) 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 likely due to the former: because a North American population expansion is not allowed until after the founding of Hawaii in the *found and grow* model, the 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 (Figure 3d). While our model optimizations span several orders of magnitude for both of these parameters, a substantial portion of the iterations settled on introductions less than 200 years ago. Since the earliest historical records of monarchs on Hawaii date to roughly 200 years ago, we are inclined to believe the results of these iterations over those with longer estimated divergence times. That the presence of milkweed (*Asclepias*, *Gomphocarpus*, *Calotropis*) host plants on Hawaii is almost certainly due recent human introductions into the Pacific also suggests a recent monarch establishment. Finally, recent sequencing of insectary-reared monarch populations has shown that as little as 20 years of captive-breeding (A. Tenger-Trolander, pers. comm.) is sufficient to generate patterns of genetic divergence recapitulating those that we observed in Pacific populations (Tenger-Trolander et al. 2019). Given this, along with phenotypic evidence for contemporary differentiation between Pacific and North American populations (Freedman et al., in review), we suggest accepting the historical records to infer the approximate establishment timing for Pacific populations. 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 S3). 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.

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 monarchs 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 small number of founding individuals reached Hawaii and then rapidly expanded their numbers in an area with abundant host plants and naïve predators. 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. We also note that a single gravid female monarch may lay an average of more than 700 eggs in her lifetime (Oberhauser 1997), potentially mitigating some of the strong post-establishment genetic drift that normally accompanies population bottlenecks.

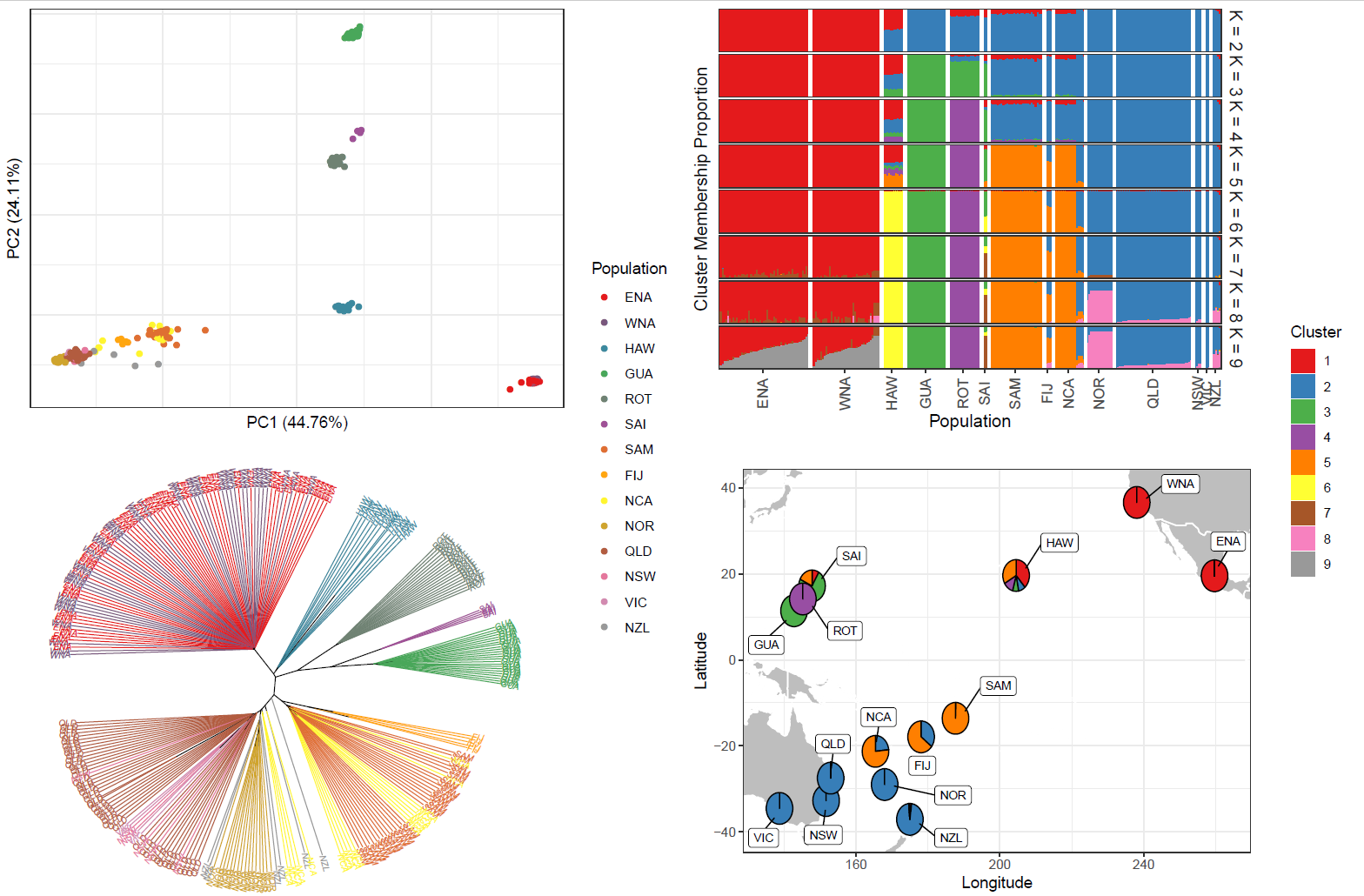
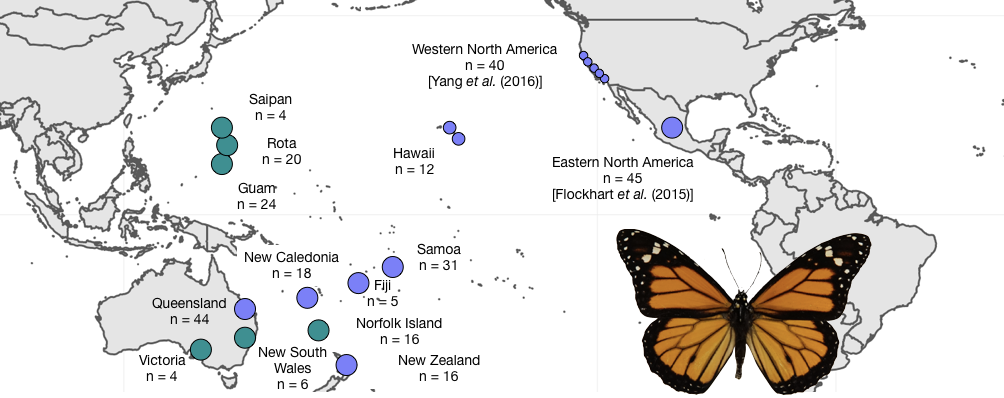
potentially mitigating some of the bottleneck effects associated with initial establishment (any references from other systems to support this?). Our model results do provide some support for this, since the model iterations with recent establishments also tended to have smaller establishment population sizes.

In contrast to variable estimates of establishment timing and founding population size, demographic models were very consistent in suggesting very low contemporary migration rates (on the order of 0.00001 individuals per generation from NA to Hawaii and vice versa). Our results thus contrast with those of Pierce et al (2014a), whose methods suggested much higher migration rates (nearly 10 individuals/generations) for both North America to Hawaii and vice versa. We are more confident in our results than those of Pierce et al. (2014) 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 -> Hawaii establishment.

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 (cite petition, Nail et al. 2019).

**References**

1. Ackery, P.R. & Vane-Wright, R.I. (1984). *Milkweed Butterflies: Their Cladistics and Biology*. Cornell University Press, Ithaca, NY, USA.
2. Ali, O.A., O’Rourke, S.M., Amish, S.J., Meek, M.H., Luikart, G., Jeffres, C., *et al.* (2016). RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. *Genetics*, 202, 389–400.
3. Als, T.D., Hansen, M.M., Maes, G.E., Castonguay, M., Riemann, L., Aarestrup, K., *et al.* (2011). All roads lead to home: panmixia of European eel in the Sargasso Sea. *Mol. Ecol.*, 20, 1333–1346.
4. Bettin, O., Cornejo, C., Edwards, P.J. & Holderegger, R. (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.
5. Brower, A.V.Z. & Boyce, T.M. (1991). Mitochondrial DNA variation in monarch butterflies. *Evolution*, 45, 1281–1286.
6. Clarke, A.R. & Zalucki, M.P. (2004). Monarchs in Australia: on the winds of a storm? *Biol. Invasions*, 6, 123–127.
7. Dawe, K.L. & Boutin, S. (2016). Climate change is the primary driver of white-tailed deer (*Odocoileus virginianus*) range expansion at the northern extent of its range; land use is secondary. *Ecol. Evol.*, 6, 6435–6451.
8. Davis, M.B. & Shaw, R.G. (2001). Range shifts and adaptive responses to Quaternary climate change. *Science*, 292, 673–679.
9. Francis, R.M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Mol. Ecol. Resour.*, 17, 27–32.
10. Freedman, M.G., Dingle, H., Tabuloc, C.A., Chiu, J.C., Yang, L.H. & Zalucki, M.P. (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.
11. Gillespie, R.G., Baldwin, B.G., Waters, J.M., Fraser, C.I., Nikula, R. & Roderick, G.K. (2012). Long-distance dispersal: a framework for hypothesis testing. *Trends Ecol. Evol.*, 27, 47–56.
12. Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H. & Bustamante, C.D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.*, 5, e1000695.
13. Helmus, M.R., Mahler, D.L. & Losos, J.B. (2014). Island biogeography of the Anthropocene. *Nature*, 513, 543.
14. Henn, B.M., Cavalli-Sforza, L.L. & Feldman, M.W. (2012). The great human expansion. *Proc. Natl. Acad. Sci. U. S. A.*, 109, 17758–17764.
15. 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.
16. Hewitt, G.M. (1999). Post-glacial re-colonization of European biota. *Biol. J. Linn. Soc. Lond.*, 68, 87–112.
17. Hulme, P.E., Bacher, S., Kenis, M., Klotz, S., Kühn, I., Minchin, D., *et al.* (2009). Grasping at the routes of biological invasions: a framework for integrating pathways into policy. *J. Appl. Ecol.*, 403–414.
18. Ibrahim, K.M., Nichols, R.A. & Hewitt, G.M. (1996). Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* , 77, 282–291.
19. Jakobsson, M. & Rosenberg, N.A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801–1806.
20. James, D.G. (1993). Migration biology of the monarch butterfly in Australia. in Malcolm, S.B. & Zalucki, M.P. (eds). *Biology and Conservation of the Monarch Butterfly*, pp. 189-200, Los Angeles Museum of Natural History.
21. Korneliussen, T.S., Albrechtsen, A. & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, 15, 356.
22. Kraus, R.H.S., van Hooft, P., Megens, H.-J., Tsvey, A., Fokin, S.Y., Ydenberg, R.C., *et al.* (2013). Global lack of flyway structure in a cosmopolitan bird revealed by a genome wide survey of single nucleotide polymorphisms. *Mol. Ecol.*, 22, 41–55.
23. Kuchta, S.R. & Tan, A.-M. (2005). Isolation by distance and post-glacial range expansion in the rough-skinned newt, *Taricha granulosa*. *Mol. Ecol.*, 14, 225–244.
24. Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754–1760.
25. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078–2079.
26. 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. *Mol. Ecol.*, 21, 3433–3444.
27. Nail, K.R., Drizd, L. & Voorhies, K.J. (2019). Butterflies Across the Globe: A Synthesis of the Current Status and Characteristics of Monarch (Danaus plexippus) Populations Worldwide. *Frontiers in Ecology and Evolution*, 7, 362.
28. Oberhauser, K.S. (1997). Fecundity, lifespan and egg mass in butterflies: effects of male‐derived nutrients and female size. *Funct. Ecol.*, 11, 166–175.
29. Paradis, E. & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35, 526–528.
30. Parmesan, C. & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37–42.
31. Peel, A.J., Sargan, D.R., Baker, K.S., Hayman, D.T.S., Barr, J.A., Crameri, G., *et al.* (2013). Continent-wide panmixia of an African fruit bat facilitates transmission of potentially zoonotic viruses. *Nat. Commun.*, 4, 2770.
32. Peter, B.M. & Slatkin, M. (2013). Detecting range expansions from genetic data. *Evolution*, 67, 3274–3289.
33. Peter, B.M. & Slatkin, M. (2015). The effective founder effect in a spatially expanding population. *Evolution*, 69, 721–734.
34. Pfeiler, E., Nazario-Yepiz, N.O., Pérez-Gálvez, F., Chávez-Mora, C.A., Laclette, M.R.L., Rendón-Salinas, E., *et al.* (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.
35. Pierce, A.A., de Roode, J.C., Altizer, S. & Bartel, R.A. (2014b). Extreme Heterogeneity in Parasitism Despite Low Population Genetic Structure among Monarch Butterflies Inhabiting the Hawaiian Islands. *PLoS ONE*.
36. Pierce, A.A., Zalucki, M.P., Bangura, M., Udawatta, M., Kronforst, M.R., Altizer, S., *et al.* (2014a). Serial founder effects and genetic differentiation during worldwide range expansion of monarch butterflies. *Proc. Biol. Sci.*, 281.
37. Pierce, A.A., Altizer, S., Chamberlain, N.L., Kronforst, M.R. & de Roode J.C. (2015). Unraveling the mysteries of monarch migration and global dispersal through molecular genetic techniques. in Oberhauser, K.S., Nail, K.R. & Altizer, S. (eds). *Monarchs in a Changing World: Biology and Conservation of an Iconic Butterfly*, Cornell University Press, Ithaca, NY, pp. 257-267.
38. Portik, D.M., Leaché, A.D., Rivera, D., Barej, M.F., Burger, M., Hirschfeld, M., *et al.* (2017). Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Mol. Ecol.*, 26, 5245–5263.
39. Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 4, 406–425.
40. Shephard, J.M., Hughes, J.M. & Zalucki, M.P. (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.
41. Skotte, L., Korneliussen, T.S. & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, 195, 693–702.
42. Slatkin, M. & Excoffier, L. (2012). Serial founder effects during range expansion: a spatial analog of genetic drift. *Genetics*, 191, 171–181.
43. Tenger-Trolander, A., Lu, W., Noyes, M. & Kronforst, M.R. (2019). Contemporary loss of migration in monarch butterflies. *Proc. Natl. Acad. Sci. U. S. A.*, 116, 14671–14676.
44. Tiedemann, R., Paulus, K.B., Scheer, M., Von Kistowski, K.G., Skírnisson, K., Bloch, D., *et al.* (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.
45. Vane-Wright, R.I. (1993). The Columbus hypothesis: an explanation for the dramatic 19th century range expansion of the monarch butterfly. in Zalucki, M.P. & S.B. Malcolm, (eds). *Biology and Conservation of the Monarch Butterfly*, Los Angeles County Museum of Natural History.
46. Weir, B.S. & Cockerham, C.C. (1984). Estimating F-Statistics for the analysis of population structure. *Evolution*, 38, 1358–1370.
47. Zalucki, M.P., Hughes, J.M. & Carter, P.A. (1987). Genetic variation in Danaus plexippus L.: Habitat selection or differences in activity times? *Heredity* , 59, 213–221.
48. Zalucki, M.P. & Clarke, A.R. (2004). Monarchs across the Pacific: the Columbus hypothesis revisited. *Biol. J. Linn. Soc. Lond.*, 82, 111–121.
49. Zhan, S., Merlin, C., Boore, J.L. & Reppert, S.M. (2011). The monarch butterfly genome yields insights into long-distance migration. *Cell*, 147, 1171–1185.
50. Zhan, S., Zhang, W., Niitepõld, K., Hsu, J., Haeger, J.F., Zalucki, M.P., *et al.* (2014). The genetics of monarch butterfly migration and warning colouration. *Nature*, 514, 317–321.



**Figure 1** - Map of sampling locations for monarchs included in our sequencing design. Points in purple correspond to locations previously sampled in Zhan et al. (2014) and Pierce et al. (2014a). Points in turquoise represent sampling locations with no previous genetic polymorphism data.

**Figure 2** – Relatedness among sampled population. (a) 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. (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) Map of sampled populations, with pie charts reflecting results from NGSadmix.

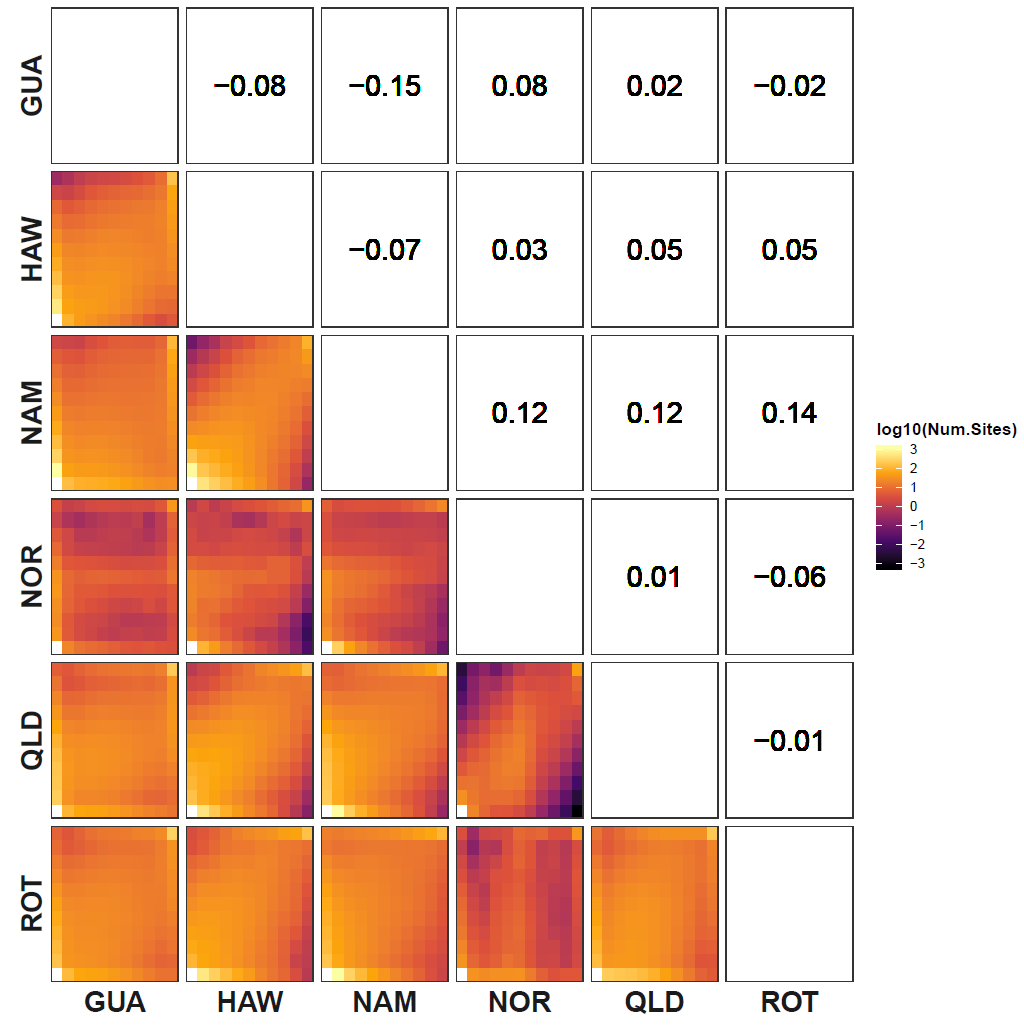


Figure 3 – Derived allele frequency spectra (below diagonal) and directionality indices (above diagonal) for each pairwise comparison between each well-sampled population. Spectra polarized via reference to putative sister taxa *Danaus erippus* and projected to 10 gene copies per population.

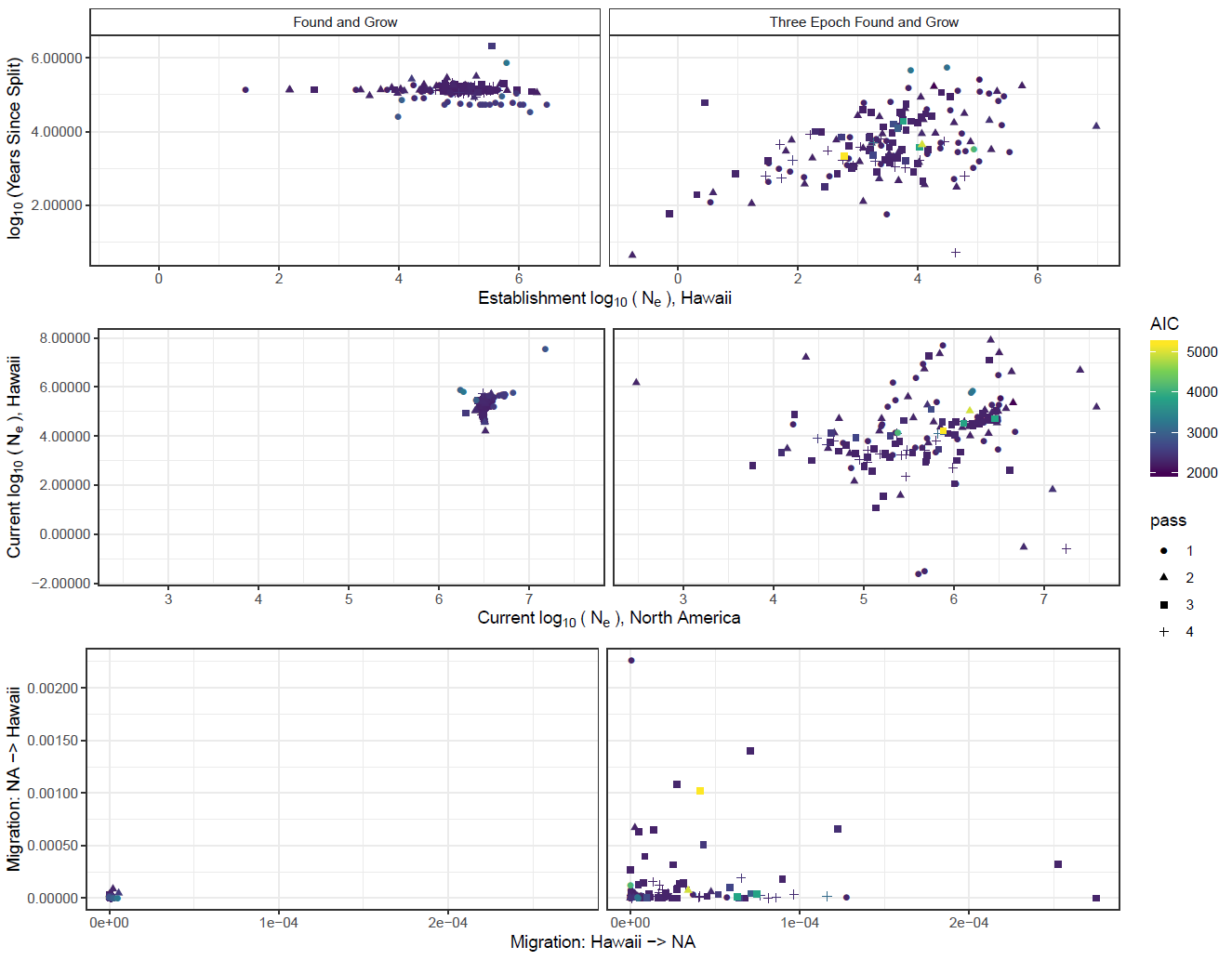


Figure 4 – Results of the dadi optimization runs for the “Found and Grow” (Left) and “Three Epoch Found and Grow” (Right) demographic models. (a) Hawaii establishment effective size and years since establishment. (b) Current effective size estimates in North America and Hawaii. (c) Migration rates from North America to Hawaii and vice versa.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Population** | **# Samples** | **Tajima's D** | **H­O** | **π** |  |
| **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’D, Observed Heterozygosity (HO), nucleotide diversity (π), and the average ratio of Heterozygous to Homozygous sites across all individuals in each population | | | | | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Population** | **NAM** | **HAW** | **GUA** | **ROT** | **SAI** | **QLD** | **NSW** |
| **HAW** | 0.026 |  |  |  |  |  |  |
| **GUA** | 0.04 | 0.164 |  |  |  |  |  |
| **ROT** | 0.037 | 0.135 | 0.164 |  |  |  |  |
| **SAI** | 0.016 | 0.146 | 0.177 | 0.227 |  |  |  |
| **QLD** | 0.036 | 0.108 | 0.177 | 0.158 | 0.204 |  |  |
| **NSW** | 0.02 | 0.107 | 0.242 | 0.208 | 0.283 | 0.038 |  |
| **VIC** | -0.025 | 0.102 | 0.28 | 0.236 | 0.377 | 0.053 | 0.078 |
| Table 2: Average pairwise FST across all sites for each pair of populations. | | | | | | | |

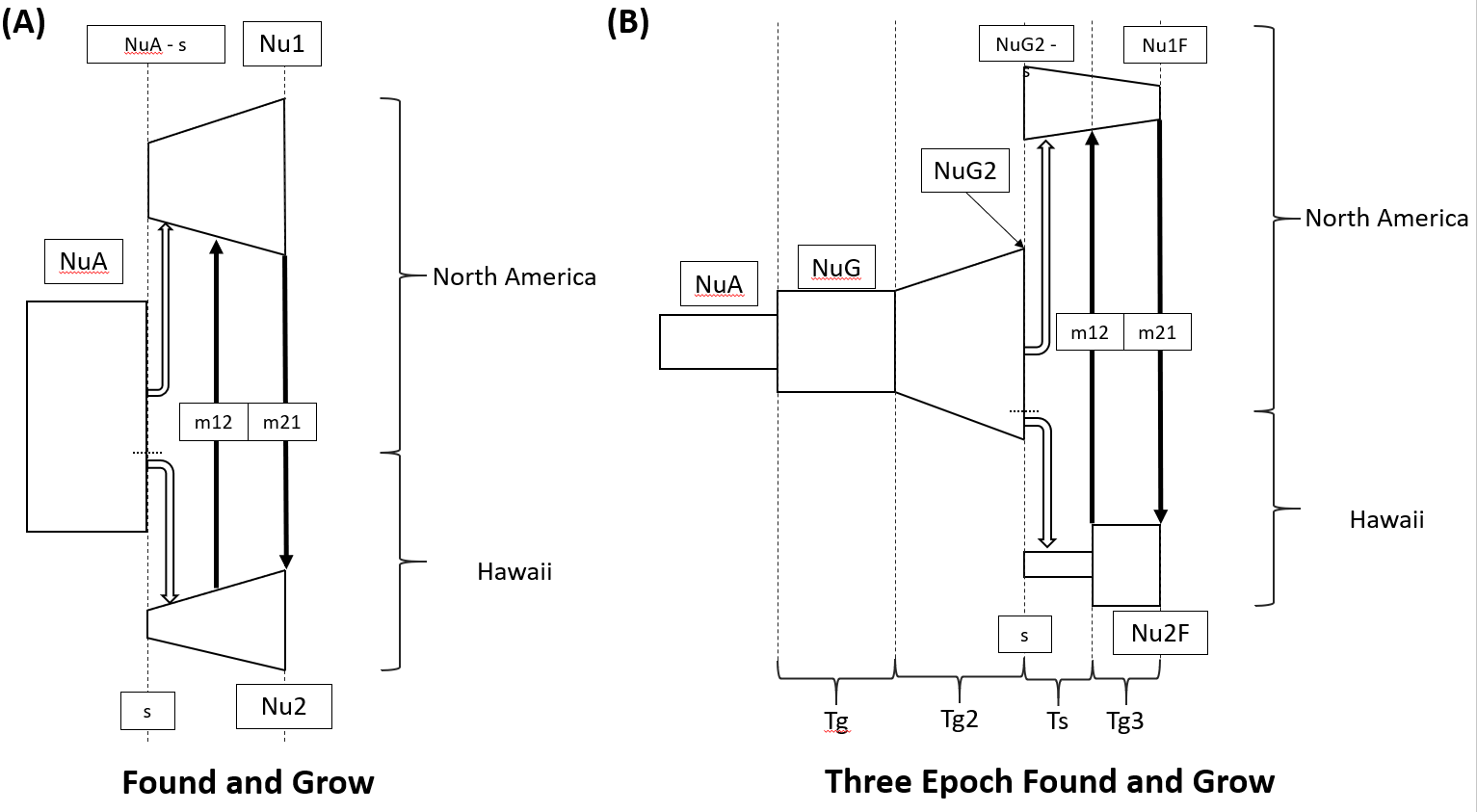


Figure S1 – Schematics of the “Found and Grow” and “Three Epoch Found and Grow” models. Note that directions of population size changes are not fixed: populations can either grow or shrink (but not go extinct) during each size change. All non-instantaneous growths are exponential.

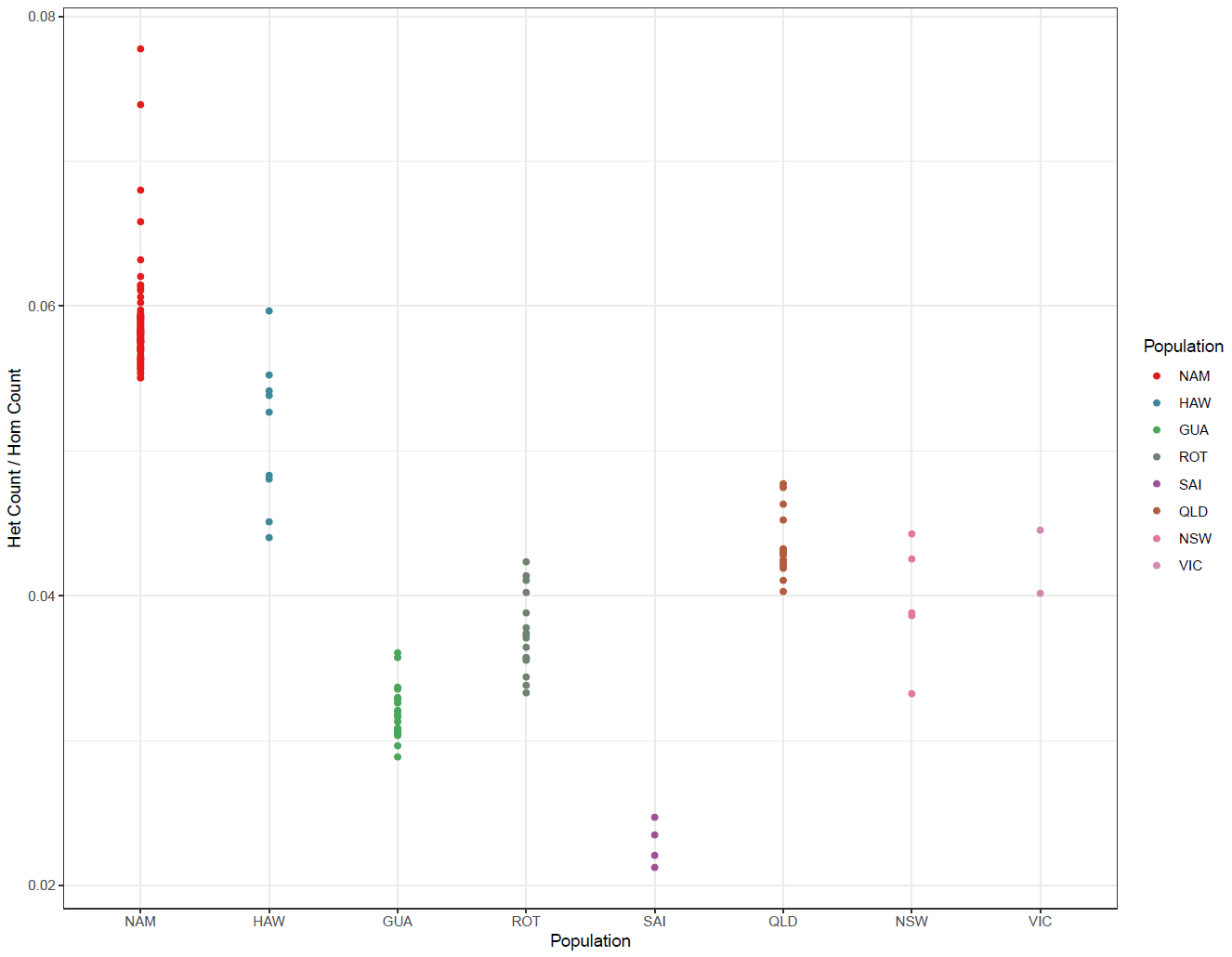


Figure S2 – Heterozygote/Homozygote ratios for each well-sequenced (>75% of possible SNPs) individual in each population.

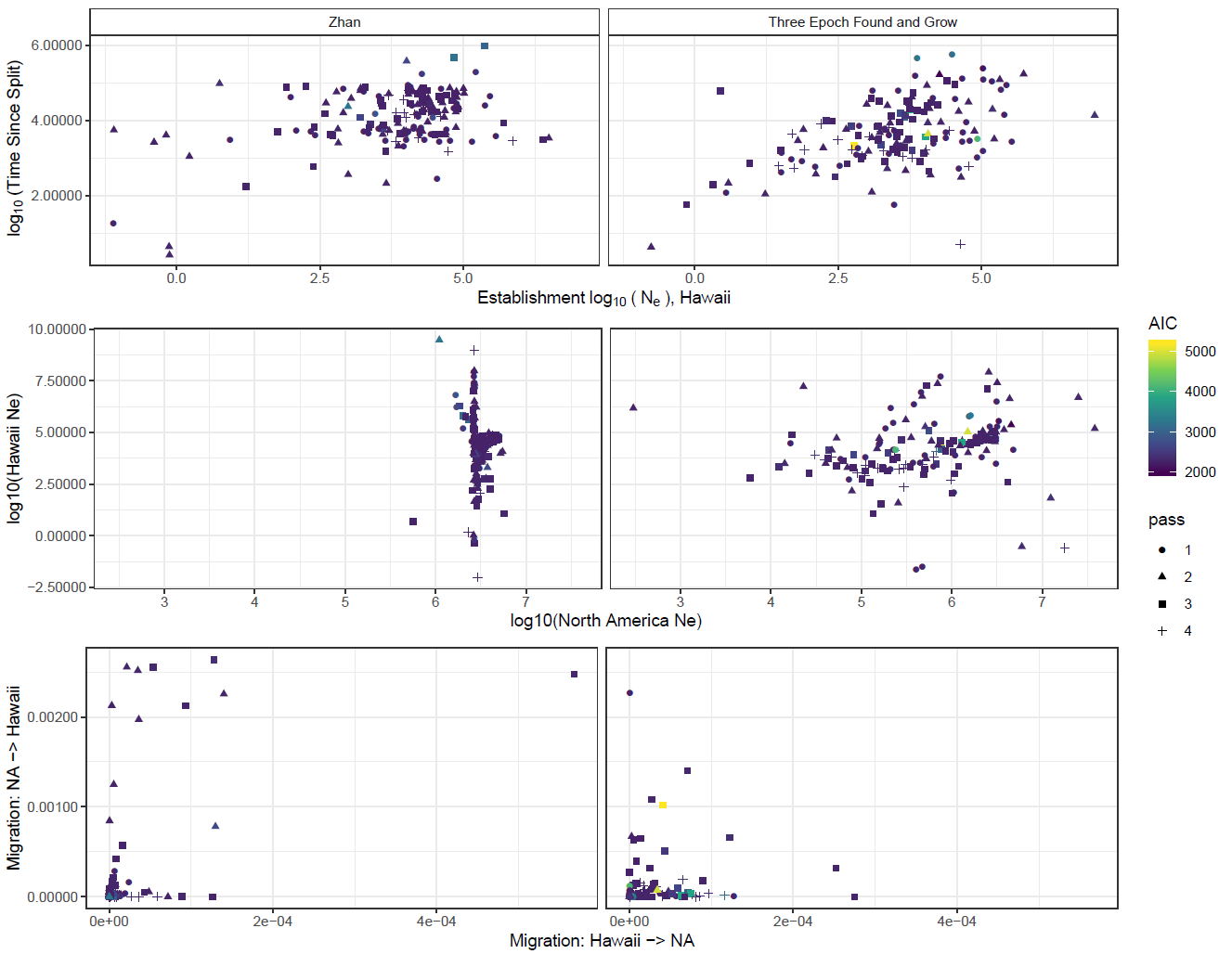


Figure S3 – Results of the dadi optimization runs for the model from Zhan et al. (2014) (Left) and “Three Epoch Found and Grow” (Right) demographic models. (a) Hawaii establishment effective size and years since establishment. (b) Current effective size estimates in North America and Hawaii. (c) Migration rates from North America to Hawaii and vice versa.

|  |  |  |  |
| --- | --- | --- | --- |
| **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. | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Run** | **Parameter Permutation** | **Number of Runs** | **Number of Optimization Iterations** |
| **1** | 3 | 50 | 30 |
| **2** | 2 | 50 | 50 |
| **3** | 2 | 60 | 50 |
| **4** | 1 | 100 | 100 |
| Table S2 – Degree of parameter permutation, number of independent dadi runs, and the number of iterations per run for each of the dadi optimization passes (see Portik et al., 2017). | | | |