Reviewer Comments to Author:  
  
Reviewer: 1  
  
Comments to the Author  
In this interesting study, the authors analyze the genetic consequences of loss of migration in populations of the monarch butterfly recently established on some Pacific islands. In my opinion, the study is well performed, addresses a question of general interest, and is nicely written (it was a pleasure to read it). So, I only have a very few minor comments that the authors might find useful to clarify/improve some parts of their MS:  
  
1) Line 98: I would change “locations” with “populations” (as only \*populations\*, not \*locations\* exchange gene flow)

**As suggested, we have updated this language.**2) Lines 131-134: Although it is obvious, I would indicate which statistics were calculated within populations (Ho, Tajima, etc.) and which ones were calculated between pairs of populations (Fst). As written, it sounds like all parameters were calculated “between pairs of populations”.

**We updated the methods section to specifically indicate that Fst was calculated between populations, while all other summary statistics were calculated within populations.**   
  
3) Lines 150-151 (Evanno´s method): I would recommend to use this method in combination with the direct interpretation of maximum likelihood values (as originally suggested by Pritchard et al. 2000). The problems with Delta K probably also apply to NGSadmix (see Janes et al. 2017). Anyway, the best solution is presenting the results for all K-values, as the authors actually did in Figure 1B (see Gilbert et al. 2012; Janes et al. 2017).

**Thank you for bringing up these important issues. We note that we still use the delta K method described by Evanno, but we updated the text to reflect the concerns that you raised here, and also to reflect that we visually examined patterns of clustering across a range of K values.**  
  
4) Line 177: Consider changing “minimum” with “earlier”.

**This section was removed.**5) Figure S1: Given the importance of visualizing the tested models, I would highly recommend to move Figure S1 to the main document (probably Figure 3 could be moved to Supporting Information).

**As suggested, the figure that was formerly Figure S1 (showing the four demographic scenarios that are the main focus of the manuscript) has been moved to the primary text and is now Figure 1.**

6) Lines 223-235 should be moved to the Results section.

**This section has been moved to the results section under the heading “Timing of establishment and patterns of ongoing gene flow.”**  
  
7) Line 256: I would suggest to change “hybrid ancestry” with “admixed ancestry”, as you are dealing with populations within a single species.

**As per your suggestion below in (8), we updated the text of this sentence to read “At K = 9, nearly all North American samples were assigned to two genetic clusters with ancestry proportions unrelated to their geographical sampling locations, which can be interpreted as the presence of a fictive cluster with no biological significance (e.g. Chen et al., 2005; Guillot et al., 2007).”**  
  
8) Lines 255-257: This is a relatively frequent outcome of clustering analyses (e.g., Chen et al., 2007; Guillot et al., 2005). However, “nonsense split” sounds a little bit weird to me. May you could change the sentence to “At K = 9, nearly all North American samples were assigned to two different genetic clusters with ancestry proportions unrelated with the geographical location of populations, which can be interpreted as the presence of a fictive cluster with no biological significance (e.g., Chen et al., 2007; Guillot et al., 2005)”.

**As suggested, we adopted the language that you suggest here and added citations to Chen et al. (2007) and Guillot et al. (2005).**  
  
9) Lines 258-259: Why delta k could not be calculated with low likelihood variance?

**The difference in the likelihood at K = 2 and K = 3 was nearly 0 and therefore could not be used in the denominator to estimate deltaK. We have updated the text to read “K = 2 and K = 5 had the highest** **ΔK values (Evanno et al., 2005), although we were not able to estimate ΔK for K = 3 due to very low likelihood variance between runs at K = 2 and K = 3, thereby producing an undefined ΔK (Figure S2).”**  
  
10) Lines 273-275: Present r-values together with p-values.

**We have added corresponding r-values for all IBD comparisons.**  
  
11) Line 340 (and through the MS): I find a little bit confusing to use “eastern and western North America” to refer to Mexican and Californian populations. Would not be easier to refer to them as Californian and Mexican populations?

**The designation of eastern versus western North American monarchs reflects the historical terminology used by monarch butterfly researchers and refers to the two main migratory regions where North American monarchs occur. Their respective overwintering locations are indeed in Mexico (eastern North American monarchs) and California (western North American monarchs), although these overwintering monarchs may have originated from anywhere within their respective summer breeding ranges. Although eastern and western North American monarchs have traditionally been treated as distinct populations based on their distinct overwintering destinations, recent research (including the data described here) shows that they are genetically indistinguishable from one another. We would be happy to add some of these details to the manuscript if you feel that this would be useful for readers.**

12) Table 1: It is confusing that Table 1 presents nine populations but Figure 1A shows 15 populations. I would recommend to include a Table presenting all populations, with their full names (not only codes). Then, you could explain that some of them were grouped for some specific analyses.

**We agree that this was confusing and have added a “Datasets” subsection to the methods to more clearly describe which populations were used for which analyses and why. Table 1 only shows nine populations because four of our populations were comprised of individuals with relatively few sequenced loci (Samoa, New Caledonia, Fiji, New Zealand). These samples came from older butterfly tissue whose DNA was likely somewhat degraded, hence the lower sequencing efficiency. The remaining two populations shown in what is now Figure 2 are the result of splitting the North American samples into east and west and the Hawaiian samples into Oahu and Maui (as per the request of previous reviewers).** 13) Figure 1: Resolution of the map is very low (pop codes are very difficult to read), some codes are missing from the legend in the PCA (e.g., for Fiji), and I think that there is not a good correspondence between the colors used in the PCA and NGSadmix plots (also purple for ROT and WNA are very difficult to distinguish in the PCA). The authors could also pay attention to other small details in the MS (software names in small caps, scientific names in italics, etc.)

**Apologies for the poor resolution. We have updated all of the figures and also added them separately as PDFs with high resolution. The PCA and neighbor joining tree have been moved to the supplemental figures, and their legends have been updated. Due to the large number of populations included (n = 15) and the choice of a color palette that is suitable for colorblind viewers, gradations between some colors are necessarily somewhat faint. We have updated the caption of this figure to point out that the purple color corresponding to western North America broadly overlaps with the red eastern North American samples, while the purple color corresponding to Rota clusters with the other Mariana Island samples. We also updated software names and scientific names so that they are formatted appropriately.**  
References  
  
Chen, C., Durand, E., Forbes, F., & François, O. (2007). Bayesian clustering algorithms ascertaining spatial population structure: A new computer program and a comparison study. Molecular Ecology Notes, 7(5), 747–756. [doi.org/10.1111/j.1471-8286.2007.01769.x](http://doi.org/10.1111/j.1471-8286.2007.01769.x)  
Gilbert, K. J., Andrew, R. L., Bock, D. G., Franklin, M. T., Kane, N. C., Moore, J. S., . . . Vines, T. H. (2012). Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program STRUCTURE. Molecular Ecology, 21(20), 4925-4930. doi:10.1111/j.1365-294X.2012.05754.x  
Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., & Andrew, R. L. (2017). The K=2 conundrum. Molecular Ecology, 26(14), 3594-3602. doi:10.1111/mec.14187  
Guillot, G., Estoup, A., Mortier, F., & Cosson, J. F. (2005). A spatial statistical model for landscape genetics. Genetics, 170(3), 1261–1280. <https://doi.org/10.1534/genetics.104.033803>  
Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2), 945-959.  
  
  
Reviewer: 2  
  
Comments to the Author  
This paper uses high resolution SNP data from RAD sequencing data to reconstruct the range expansion pathways of Monarch butterflies from North America across the Pacific, including Australia. The story is of broad interest and the conclusions are compelling and the paper clearly written. They document a scenario of step wise dispersal across the Pacific with support for a single expansion. The paper is a great contribution to the natural history of Monarchs but also in understanding global insect expansions. The methods used are appropriate and make good use of some powerful modelling approaches.  
  
Abstract –   
Line 10: So is it 284 butterflies from 15 ‘pacific locations’ in North America? What is a pacific location? A bit confusing- be more specific.

**We have updated the text of the abstract to read “Here, we present reduced-representation sequencing data for a total of 281 monarchs collected from either North America, one of 12 Pacific Islands, or from three locations in Australia.” The 15 “Pacific locations” refer to 12 individual Pacific Islands (some of which are islands within single archipelagos) plus three locations within Australia.**  
  
Line 126 – it isn’t really clear why strong bottlenecks would be expected in the samples, and how this really justifies not having a MAF which is important for dealing with sequencing errors.

**We expected strong bottlenecks because of previous research showing stepwise colonization patterns in some of the Pacific locations that we sampled, and we are assuming that each island was colonized by a relatively small number of individual butterflies. We have updated the text of this section to read “Since strong bottlenecks are likely to cause large differences in allele frequencies between populations, which, in conjunction with very different sample sizes between populations, can result in loci with very low overall minor allele frequencies having relatively high frequencies in individual populations, we did not use a minor allele frequency filter when calling genotypes.”**There is no indication of how Hardy Weinberg Equilibrium deviations were accounted for and there is no indication of how linkage disequilibrium was accounted for – the data should have been filtered for these aspects prior to analysis of genetic structure.LD is often minimised also by creating a data set of one SNP per RAD tag but this is also not performed. Without doing this it is impossible to know how much your data could be biased by HWE and LD.

**We have updated our methods to address the suggestions you made here. To address potential issues arising from LD between loci, we randomly subsampled SNPs within 10,000 bp of one another. This filter was applied to all analyses of demographic history as well as calculations of most summary statistics (π, Ho, Fst, IBD relationships). We also tested for SNPs that violated HWE assumptions; when applying the methods of Wigginton et al. (2005), these comprised a relatively small fraction of the total number of called SNPs (0.76%, or 86 out of 11,384 loci) and had only a minor impact on our inferences. Thus, for our primary analyses, we continue to report results without this HWE filter, although we do report the results of the same analyses *with* the HWE filter in the supplementary materials (Tables S5 and S6).**Please consider using more subheadings in your methods section to describe the different types of analyses done and their purpose. It feels like the methods are jumping around between approaches and objectives and they are not entirely logical. Subheadings would help a lot. Link your aims to your methods more clearly.

**As suggested, we have updated the methods section so that it is now arranged into the following sections and subsections:**

* **Sample preparation and sequencing**
* **Sequence alignment, filtering, and genotype calling**
  + **Datasets**
* **Patterns of relatedness among monarch populations**
  + **Serial expansion**
  + **Genetic variation across space in migratory vs. non-migratory populations**
* **Demographic history of the monarch’s expansion**
  + **dadi model selection**
  + **Parameter estimation**

Line 164 – IBD relationships do not tell you really about migration. If you want to look at Migration use GeneClass 2 or Migrate. Rephrase or add more analyses to test for migration rates explicitly.

**As suggested, we have updated this section to read “We looked for evidence of isolation by distance (IBD) between samples from the Mariana Islands (non-migratory), Hawaii (non-migratory), Australia (partially migratory), and North America (migratory).” Here, our approach is not to explicitly analyze rates of migration within or between populations, but rather to coarsely characterize how IBD relationships differ between monarch populations with different propensities to seasonally migrate.**Line 149 – what do you mean ‘to help simplify’ the analysis?

**We have updated this language so that it now reads “The pophelper (Francis 2017) and snpR (Hemstrom and Jones 2021) R packages were used to run these analyses.”**

Line 155 – please give a bit more information about SFS – how it works, what it is etc. The approach is not that widely used (yet).

**We have added the following text to hopefully clarify what SFS-based demographic approaches involve: “Briefly, since demographic processes influence the frequency of common or rare alleles across loci, and an SFS describes how many individual loci fall into each possible allele rarity in each population, a SFS can be used to infer historic population processes. dadi therefore uses simulation to compare the SFS predicted under a specific demographic history to the SFS observed from the data in order to evaluate the likelihood of a demographic model and optimize the parameters of that model.”**Line 223 – it isn’t clear what is meant by ‘found and grow ‘ and two epoch – there needs to be a clear brief description of what these models involve when they are mentioned. The names are far form intuitive. Further there are results reported here – can these be moved to the results? lines 234-235 are not necessary.

**We have added a new figure to the main text (Figure 1) that shows graphical representations of each of the demographic scenarios outlined in the methods. The text from lines 234-235 has been removed entirely from the manuscript.** 299- have you considered not presenting all three models in the main text if the models are so similar? Could you choose the best performing one? The SFS figure is rather homogenous looking with all three models.

**We have updated what was previously Figure 3, which showed site frequency spectra for all of the focal demographic models, so that it now only focuses on the *Three Epoch* model that we discuss most thoroughly. However, we retained parts of the methods and discussion that describe the other candidate models (*Two Epoch, Found and Grow, Zhan*), in part because they had similar overall likelihoods but fairly divergent parameter estimates. We recognize that this may make the interpretation of the results and discussion of the manuscript somewhat tedious, although this is also a result in and of itself: there are multiple demographic scenarios that generate similar observed site frequency spectra, and researchers should use discretion and their knowledge of the biology of the system to interpret results accordingly when performing similar analyses.**Line 306 – this is a statement that belongs in the Discussion. It is an interpretation, not a result.

**This statement has been removed from the methods section.**Line 374 -376 – this is in contrast to what is said in the results it seems ie..that the models were similar. Reading this, perhaps presentation of the Three Epoch model alone is sufficient for the main text.

**We have updated the text of the results to expand more on the different interpretations produced by each demographic model (lines 329-343), particularly with respect to establishment timing. We still report on parameter estimates that are concordant between models. Our discussion also continues to feature a brief mention of the four competing models, but we focus nearly all of our attention on the Three Epoch model.**It would be good to hear more about the conservation status of Monarchs and any predictions related to climate change and their distribution, and how your results are informative in relation to these issues.

**We rewrote the last paragraph of the discussion to focus more on the issues that you bring up here. Monarchs were recently considered for listing under the U.S. Endangered Species Act, whose decision was that a threatened listing for monarchs is “warranted but precluded.” This decision effectively says that the population trajectory of monarchs and their ongoing decline in North America is grounds for a listing, but in practice, the U.S. Fish and Wildlife Service does not currently have the capacity to enforce this listing. We added a brief mention of how climate change is expected to influence patterns of seasonal migration in monarchs, although we are not aware of any studies that attempt to estimate how it might contribute to expansion (or contraction) in their global distribution. Finally, we added a few sentences discussing how our results are informative broadly for monarch conservation.** Table 1 – please indicate geographically where the sites are in the legend.

**We have updated Table 1 to include full location names and also added the following text to the table legend: Populations from Guam, Rota, and Saipan are all part of the Mariana Islands archipelago. Queensland, New South Wales, and Victoria are all within the Australian continent.**Figure 1 – This figure was at very low resolution in my copy – it appears that it is also too busy and I would consider dividing the figure into two or excluding one of the panels to enlarge them all. The round ‘dendrogram’ looking plot also has not labels and it is not clear if this is needed. I think at least one of these panels can go in supplementary material. They all kind of say the same thing.

**Apologies for the low figure resolution; it should be better now. As per your suggestion, we have split what was Figure 1 into two separate figures. The figure retained as part of the main text only shows the NGSadmix results and the map of sampling locations, while the PCA and neighbor joining tree have been moved to the supplemental materials.**Figure 3 – the axis labels on these plots and the legend need to be more informative- same with in Figure 2. The plots look very similar and I wonder if all of them need to be presented – if there are ones statistically more differentiated then perhaps rethink the presentation in the main results here. The legend is not clear enough to understand the figure on its own.

**We have substantially updated what was Figure 3 (now Figure 4) to hopefully address some of the issues you raise here. We agree that the plots looked quite similar and that it was difficult to distinguish meaningful differences between the site frequency spectra generated from each candidate model. As such, we now use Figure 4 to present a comparison of the dadi model-estimated versus observed site frequency spectrum for only the demographic model (*Three Epoch*) that is the primary focus of our discussion. We also edited the caption to what is now Figure 3 (previously Figure 2), which shows parameter estimates for dadi optimization runs, to more carefully explain the legend and what it denotes. Please let us know if you still find the presentation of the figures to be confusing and if you have suggestions for how to more clearly summarize these findings.**Table S4 seems to be missing from the supp material.

**Apologies, Table S4 was quite large and was not included as part of the supplemental material. We have included it as a .xls file with this submission.**  
  
Reviewer: 3  
  
**Comments to the Author**<b>Review: </b>  
In this paper the authors have used RAD sequencing data from 281 monarchs from North America and 15 locations across the Pacific to analyze their demographic history using dadi and dispersal patterns using population genetic statistics.  They have used the models “Found and Grow”, “two epoch”, “three epoch” and a variation from <i>Zhan et al. 2014 </i> to determine the patterns of monarch pacific expansion.   
  
The authors addressed most of the issues raised by the reviewers in their previous submission to Proceedings b.  I agree with the stepwise dispersal of the monarchs and it is supported by their FST analysis, but I have a few issues with this paper that were not raised by the previous reviewers.   
  
<b>Major issue / confusion: </b>  
Why is the nucleotide diversity (π) in these small island populations so high? In Table 1 all estimates of genetic diversity are >0.02 in these island populations and >0.06 in North America. While the genetic diversity given in <i>Talla et al. 2020</i> is ~0.011 in eastern and western North America. The highest genetic diversity in butterfly species is ~0.042 (in 4fold sites) reported in (Martin et al. 2020).  Why do these island populations have such high genetic diversity? Is this a different statistic? or is it provided in percentages?

**We agree that the estimates of π reported here seem high relative to other published papers on monarchs and other butterflies. However, it is difficult to directly compare across studies because of differences in sequence generation, sequencing coverage and depth, and especially filtering steps. The estimates of genetic diversity that we report are best understood relative to one another, rather than in direct comparison with other published papers. A major difference between our estimate of π and that reported in the papers that you mention is that we only considered polymorphic loci when calculating the denominator, and did not take include sites that were invariant across samples. We have clarified this in the caption of Table 1 and could redo the calculation of π to include invariant sites if you feel this is warranted.**

Even the Het/Hom values don’t compare with <i>Zhan et al. 2014</i>. I think you have too low values here.

**Thank you for pointing this out. The Het/Hom ratio that we initially reported did differ substantially from that of Zhan et al., who found values generally in the range of 1-2. The major discrepancy here is that their method likely (although it is not stated explicitly) was based on calculating the ratio of heterozygous sites to *non-reference* homozygous sites, whereas the value we reported was comparing the ratio of all heterozygous sites to all homozygous sites. To avoid confusion, and because the information provided by this metric is largely redundant since we also report π and HO, we have removed the Het/Hom ratio from Table 1 and elsewhere in the paper.**

Even after considering these diversity numbers relative between these groups. Why is the NAM (π) so close to the HAW (π). We know for a fact that the North American monarch population size is many folds higher than Hawaii.

**While it is certainly true that the present census population size within North America is much larger than that in Hawaii, the effective population size within North America has been shaped by (1) its demographic history, which includes a recent population expansion and large year-to-year variation in population size and (2) its seasonal migration, in which the entire North American population is “founded” each year by the relatively small number of individuals that survive overwintering and recolonize the breeding range. By contrast, Hawaiian monarchs occur in a much less seasonal environment and are therefore subject to less demographic stochasticity. Thus, while Ne in North America (suggested by Zhan et al. to be approximately 2 million individuals) is probably at least two orders of magnitude lower than the census population size, in Hawaii these values are probably more concordant. Still, we agree that the comparatively small reduction in π in Hawaii vs. North America is somewhat surprising and probably also reflects rapid population expansion upon establishment in Hawaii.**

I’ve read in the past the RAD-seq data underestimates genetic diversity, but this seems like the opposite pattern. Please explain if I am missing something here.

**Yes, you are correct that RAD-seq data is generally thought to lead to underestimates of genetic diversity relative to other sequencing methods. This reduced diversity likely arises from biases in restriction enzyme cut sites (i.e., cut site polymorphisms can lead to unequal sequencing across loci and allele dropout), and the pattern is more pronounced in species with high levels of polymorphism. As we mention above, the estimates for π and Ho that we present may appear to be very high, although this is mainly due to the fact that we only included polymorphic loci in our calculations.**

<b>Github doesn’t work</b>  
I tried to find your scripts to verify my concerns but the link to Github doesn’t work and I couldn’t find the data.

**Apologies, at the time of submission, the repository containing all analyses was still set to private. It should now be publicly accessible at this link:** [**https://github.com/hemstrow/F-H\_2018**](https://github.com/hemstrow/F-H_2018)

<b>Discrepancies between models:</b>The authors don’t confidently point out to the best model that fit the data with the lowest AIC. The AIC values seems to fall too close to each other to confidently eliminate the other. Three epoch model  seems to indicate a much border founding population size in Hawaii. I am thinking that it is caused by the high genetic diversity in the HAW and other island populations.

**As you suggest, it is difficult to identify a single best demographic scenario using AIC values alone. Hence, we chose to report the results of all of the top-performing models and focused our discussion on one of the models (three epoch), since it produced parameter estimates that best conformed to our prior understanding of the timing of the monarch’s range expansion. For the sake of clarity, we chose not to fully describe the outcomes of the other top-performing models, in part because we have strong evidence for an upper bound on the timing of establishment that falls outside of the estimates produced by these models. Monarchs require milkweed host plants to complete their life cycle. Because their host plants in most Pacific locations are themselves recently introduced, the earliest possible establishment dates would be ~3,000 years ago (coincident with the expansion of Polynesian people), or more likely ~150-200 years ago (coincident with the onset of global trade networks that included Pacific islands).**

**We note that the three epoch model produced a broader range of outcomes for all of the estimated parameters (establishment timing, founding population size, current Ne). The high genetic diversity observed in Hawaii could plausibly be associated with a recent introduction by a small number of individuals (Figure 3, bottom left quadrant of the top left panel); in this scenario, not enough time may have passed for drift to cause the loss of rare North American alleles present in the Hawaiian founding individual(s).** I am willing to look at the MS again if these concerns are properly addressed.