

Review of: "Genomic signatures of isolation, hybridization, and selection during speciation of island finches"

Thomas Blankers¹

¹ University of Amsterdam

Potential competing interests: The author(s) declared that no potential competing interests exist.

This is a nice paper although it is very densely packed with difficult analyses some of which are would benefit from additional details. Also, it is a little bit unclear what alternative speciation scenarios are considered. The main question of the paper seems to be: "How did the Sao Tome grosbeak and Principe seedeater speciate?". But the only scenarios that are considered are 1) divergence with primary gene flow and 2) divergence in isolation followed by secondary contact. I see no consideration of the scenarios where the grosbeak and seedeater independently colonized the islands (rather than speciate in situ) and the grosbeak is disappearing and there is also no discussion about whether these guys are species at all or whether there is a polymorphism in the population and one of the types is just rare/disappearing. The latter point may not be likely, but there does seem to be a critical assumption the authors are making (these are species and they are still diverging) that impacts how they interpret their analyses, specifically in relation to regions related to bill morphology protected from gene flow by selection. An important detail is for example that reproductive isolation is now mentioned only once in the discussion:

"The substantial size difference between the grosbeak and the seedeater would constitute an effective reproductive barrier important for completing the speciation process"

When I read this sentence, it sounds like there is no evidence for speciation beyond an ecological polymorphism, although I am aware of similar situations in other bird systems where strong reproductive isolation seems to exist between only minimally diverged populations.

Then I have some concerns about the analyses:

1. I like the allele sharing analysis because it makes a lot of intuitive sense. However, I am concerned about the effect of the sample sizes and the demographic history (see remark 2 below). First, it seems like a relevant detail that you only sampled 4 grosbeaks. This detail became only clear to me after going through supplementary material with sample listings. This info needs to be more readily available in the text/figures. Second, how will this bias the allele sharing analysis? You mention some subsetting for this analysis in the stacks pipeline section, but this information is not enough to understand what you did. Also, you mention in that sentence in the methods both a full data set analysis and a subset of 4 for all populations, but you do not specify to which analysis the results shown in the main text correspond.
2. What about bottlenecks/population reductions? If the grosbeak is critically endangered as you mention in the

introduction and recently settled (split time < 1 M years ago and evolved in isolation) then this should have strong signatures on the genome either on nucleotide diversity, excess homozygosity (Tajima's D) or both. Did you explore the possibility for a bottleneck/population size reduction? And what would that do with numbers of private alleles? For example, wouldn't strong drift effects in small populations increase the number of private alleles by chance? And if there is no evidence for population contraction or bottlenecks, then is it still likely that grosbeaks and seedeaters speciated in situ on Sao Tome and then came into secondary contact?

3. I am puzzled as to how the abba-baba test was performed. To me it seems like you tested P1 = allopatric seedeater, P2 = sympatric seedeater, P3 = grosbeak, P4 = outgroup. In that case, a positive D statistic would support excess allele sharing between P2 and P3 and a negative D statistic would support excess allele sharing between P1 and P3. The expectation is the former and this is also what you conclude, but I was surprised to see this conclusion is based on a negative D statistic (in the suppl. table; in the main text the number is positive). I must be missing something here, but I cannot deduce it from the methods or the supplementary table.
4. The genomic segment distribution and intersection with bill morphology genes also raises some questions. First, what does it mean when a bin does not have a genomic segment? No SNPs in that bin that map to the Zebra finch genome? So, does this aspect not influence the strength of the relationship between relative distance and genomic segment type? Are "preserved" segments not simply more likely to be differentiated between Zebra finch and grosbeak and therefore have lower mapping rates in regions far away from genes? In that case, the effect you're seeing is independent of the evolutionary processes that you are connecting to segment types and instead dependent on sequence similarity. A clear formulation of the null hypothesis with some support for the validity of the null is required. Simulations of the effect of mapping rates on the outcome of the analysis would go a long way. Such simulations may also be useful as an alternative to the bootstrapping approach you use for creating a null distribution for the relationship between genomic segments and bill morphology genes. This bootstrapping seems to be a somewhat naive and unconservative approach as just throwing the segments on the genome irrespective of gene density is probably not going to give a fair null distribution of the chance of finding a segment in the vicinity of a bill morphology gene.

I think a lot of these issues can be resolved with more details and clarification, but some may require additional analyses. This study brings new insight about the divergence process of these birds, with potential implications for island radiations/divergence more generally, but the complexity of the analyses and the limited discussion of alternative evolutionary scenarios make it somewhat inaccessible in its current form.