

Review of: "High-altitude adaptation and incipient speciation in geladas"

Haibing Xie

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

In this manuscript the authors assembled the genome of the gelada and studied the genetics of incipient speciation of geographically isolated gelada populations and their adaptation to high-altitude environments. The authors revealed a centric fission of chromosome 7 that may be in an association with reproductive barrier between the geographically isolated Northern ($2n = 44$) and Central ($2n = 42$) gelada populations. The gelada is a very famous primate in Ethiopia and this study is valuable in understanding gelada's adaptation to high-altitude environments and exploring the genetics of incipient speciation process under geographical isolations. Overall, the results are very interesting and will have a broad readership. There are a few concerns/suggestions for the authors as described below:

1. In Figure 3A, the author used MSMC to infer the demographic history of two gelada populations (Northern and Central) and to speculate that the two gelada populations began to diverge about 500 thousand years ago. I do not think it is accurate to date their divergence time based on the match of population size curves. The N_e trajectories inferred by PSMC and/or MSMC can be easily affected by gene flow that distorts the N_e curves of admixed species (please find the paper: <https://pubmed.ncbi.nlm.nih.gov/32384152/>). At the same time, the N_e change is sensitive to local environmental/ecological factors, and the population split is not necessarily correlated with the N_e changes. MSMC and PSMC required higher sequencing depth, as the sequencing coverage here is $\sim 10X$. It is recommended to infer the split time for the two populations using MSMC to estimate the relative cross coelenterate rate for each population pairs, or other tools like Coal-HMM.

2. Inferring from the data of Figure 2D, the majority of Northern and Central (including Zoo) populations showed distinct genomic components. Given the large difference between the two populations, any hybridization between them should show a clear mixture signature in the offspring genomes. The three heterozygous zoo individuals ($2n = 43$) had a fraction of Northern ancestry (10-25%?), at a level much lower than an expectation if they had a direct origin from hybridization events. Otherwise if the three individuals had hybridized ancestors many generations ago, does it mean that, perhaps, the heterozygous ancestors were still fertile, or at least partially fertile? I suggest the authors to compare the success ratio of reproduction of geladas ($2n=42$ and $2n=43$) in captive to further confirm the effect of hybrid sterility. In addition, the Central population still contains some Northern ancestry, and it seems more likely the

incompatibility may involve incomplete hybrid sterility, if the two populations had any form of hybrid incompatibility. Partial hybrid incompatibility is quite common in incipient speciation.

Suggested works on partial hybrid incompatibility in incipient speciation: Xie *et al.* 2021. Genetic architecture underlying nascent speciation - The evolution of Eurasian pigs under domestication; Turner *et al.* 2012. Reduced male fertility is common but highly variable in form and severity in a natural house mouse hybrid zone.

3. Despite the observation of the centric fission of chromosome 7, the authors did not compare the genome evolution of chromosome 7 and the remaining autosomes, as well as the X chromosome. The gelada karyotype difference is reminiscent of the observation on human chromosome 2 in a comparison to closely-related primates. The chromosomal rearrangements are important in evolving reproductive barriers, but the authors still need to compare the differentiation pattern of chromosome 7 and the rest of the genome between the Northern and Central populations, using the resequencing data. According to the general understanding of incipient speciation, the genes associated with postzygotic reproductive isolations usually accumulate genetic differentiation faster and are relatively refractory to gene flow. The authors should conduct the comparison to clarify the genetics of incipient speciation of the two populations.

Suggested citation:

McConkey 2004, Orthologous numbering of great ape and human chromosomes is essential for comparative genomics.

4. Additional to the karyotype difference, the authors should integrate the gelada reference genome and resequencing data of populations to refine the structural change of chromosome 7. Is the karyotype difference associated with the deletion/duplication of any genes on chromosome 7? Where is the boundary? What's the genomic difference between the two populations that are in a tight connection with the karyotype difference? Answering these will provide key information about questions on the incipient speciation.

5. The authors used BUSCO to measure the completeness for genome assembly, and it is recommended to run BUSCO analysis for predicted protein sequences as well (--mode proteins) to provide additional information about the completeness of gene model annotation.

6. The genome assembly for gelada is only based on NGS reads, which have low ability to cover repetitive sequences and highly similar sequences(see two papers: <https://www.nature.com/articles/ng.3631>; <https://www.nature.com/articles/s41586-021-03451-0>). The comparative analysis of gene family across species highly relies on the completeness of genome assembly. As the author reported, the genome assembly only covered 91.7% and 89.0% of expected genes present and complete in mammals and

primates, which is incomplete. I suggest the authors to weaken the analysis of gene family in the manuscript.

7. It is unclear why the author used site model, but not branch-site model in PAML for detecting positive selection? Please clarify.