Review of: "High-Quality Genome Assembly of the Endemic, Threatened White-Bellied Sholakili Sholicola albiventris (Muscicapidae: Blanford, 1868) From the Shola Sky Islands, India"

Ksenia Krasheninnikova[1]

1 Wellcome Sanger Institute, Cambridge, United Kingdom

Potential competing interests: No potential competing interests to declare.

The manuscript presents a genome assembly of Sholicola albiverntris - an avian species endemic to a restricted area in India.
Authors provide a comprehensive description of the methods used for species collection, sequencing, and assembly. The presented QC metrics of data and assembled contigs meet the criteria for good quality data. This work would benefit from the kmer checks performed for the assembly, such as kmer completeness and QV score, and from the analysis of kmer plots (all described here).

It would make sense to provide information on how the data was basecalled - was it Dorado, Guppy / other basecaller, and also if it was basecalled in standard or super accurate mode?

Polishing has to be described in more detail. Six rounds of polishing can look a little excessive, at least at first glance. How was it ensured that the assembly was not overpolished?

The content of Table 1 has to be double-checked, as the value of Contig N50 (Mbp), which equals 22.52 Mb, doesn't align with the value of the Largest contig (Mbp), which equals 8.27 Mb.

It is not fully clear what is presented as the final assembly. It can be suggested that the polished contigs after haplotigs removal are presented as such. The scaffolds obtained by RagTag are inevitably biased by the *Taeniopygia guttata* genome assembly. Also, this is the reason why synteny analysis of the pseudochromomes is not useful for homology inference. If the authors want to keep the synteny checks in the manuscript, they have to be updated using contigs of the original unscaffolded assembly.
The best source to produce scaffolds would be to employ HiC or another kind of long-range sequencing data from the same species. In the absence of such data, it's worth releasing an unbiased version of the contig assembly as the primary assembly.

If there is any information about how rearranged the species of*Taeniopygia guttata* and *Sholicola albiventris* are relative to each other, it would be helpful. If there is a well-grounded expectation that the genomes are fairly conserved, then the produced reference-based scaffolded assembly can be used for a careful estimate of how many chromosomes were

assembled in how many fragments. It will also be valuable to compare it to the karyotype information of this species if such is available.

At the time of review, I wasn't able to access the entries JBDGPF000000000, JBDGPF010000000, and PRJNA1096119 on NCBI. It's probable that the data hasn't come through to become publicly available yet.

Lastly, it could be recommended that the figures are provided in high resolution, as it's not possible to effectively zoom in.