

<https://doi.org/10.1038/s42003-025-07826-1>

Reply to: The genomic structure of complex chromosomal rearrangement at the *Fm* locus in black-bone Silkie chicken



Qiang-sen Zhao, Feng Zhu & Zhuo-cheng Hou

REPLYING TO A. Sharma, *Communications Biology* <https://doi.org/10.1038/s42003-023-05619-y> (2024).

High-quality genome assembly can effectively address some of the challenges of biological evolution. Our work¹ has assembled a Silkie genome (CAU_Silkie) using a composite approach, resolving in one fell swoop the complex genomic variation of the fibromelanosis (*Fm*) trait and identifying a large number of important genes related to metabolism, immunity, and reproduction in birds. Below we respond to Nagarjun's comments on the genomic structure of the *Fm* locus, the various lines of evidence, and possible sources of error.

The structural variation involved in *Fm* includes not only genomic duplications but also inversions, especially since the individual assembled in our work is a heterozygote (*Fm/fm*). The validity of our results is supported by several solid lines of evidence. Firstly, both wild-type (**N*) and mutant-type (**Fm_1*) structures are supported by continuous long ONT reads. The ONT reads (length > 50 kb) were aligned to both **Fm_1* (CAU_Silkie, **Fm_1*) and **N* (GRCg6a **N*), retaining only complete reads (alignment length/read length ≥ 0.99) and removing secondary alignments. The results showed that the reads from this individual support the structure of both scenarios (Figure S1–2). As the genome submission, we provided only the mutant-type assembly that can represent the *Fm* phenotype. We then calculated the genome coverage (Figure S3). The depth of the ONT and HiFi reads approaches that of the whole genome, within a sequencing depth of 0.5–1.5 times the depth of the whole genome. This indicates that the structure at this locus is supported by long reads. At critical junctions, we randomly selected reads longer than 50 kb as examples for demonstration (Figure S4). At this point, the single read can prove the junctions of two components (e.g. dup1 and reversed dup2.) in short range. Secondly, the String Graph and contigs from haplotype assembly further confirm the validity of our results. The String Graph is the optimal choice for resolving segmental duplications². The String Graph (Fig. 1) clearly shows that the genome of heterozygous individuals (*Fm/fm*) has two parallel independent haplotypes in the *Fm* region. Haplotype assembly using high-quality HiFi also resulted in two complete sets of independent contigs. Contigs from assembly related to the *Fm* locus sequence show that **Fm_1* is the correct scenario (Figure S6). Another view is the clipped reads visualization from IGV³, which also shows **Fm_1* is the correct scenario for CAU_Silkie in short range (Figure S5). All the evidence suggests that our results are correct.

Nagarjun et al.⁴ mentions some 'key pieces of evidence' that argue there is something wrong with our assembly, but they are all quite problematic. Nagarjun et al.⁴ questioned the lack of read support for the **Fm_1* junction order in Matter Arising. However, our results¹ have clearly demonstrated strong support for the junction order of the *Fm* locus using the ONT reads and HiFi reads in Fig. 1b and S12–S15¹. The coverage of the junctions of each duplication is nearly identical to the average coverage of the whole genome, and each junction is covered by complete reads, meaning that the assembly is free of structural errors (indicated by the y-axis Fig. 1b¹; Figure S4).

Nagarjun's comments also argue that the position of the locus found by the Haplotype Defining Positions (HDP) method⁴ in CAU_Silkie is incorrect, leading to the conclusion that the mutation type of *Fm* is *Fm_2*. It is clear from the above results that the individual used for CAU_Silkie were genotypically heterozygous and that the Silkie genome would have had a large number of de novo mutations compared to the red jungle fowl genome (GRCg6a). However, Nagarjun's method of determining haplotypes was to align the data to the red jungle fowl genome while removing tri-allelic loci, which meant that some of the information that actually contained haplotype differences was discarded. This makes it easy to understand why some of the SNPs are mislocated on CAU_Silkie, as these loci are not mutant haplotypes at all. HDP is not accurate when applied to heterozygous individuals. Most importantly, when the reads from heterozygous individuals are aligned to **Fm_1*, Figure S7 explains why HDP can't handle the phasing job, leading to wrong phasing results and no tiling path. In fact, Nagarjun's results also contain some errors. From the data provided by Nagarjun, we added a column named 'Hap base in Silkie genome' (Supplementary Data 1), which shows the real reference base pairs of CAU_Silkie, but after manual confirmation, most of the base pairs (column named 'Hap1 base in Silkie genome') were wrong ((92 + 98 + 114 + 111)/(121 + 121 + 137 + 137)), it explains the result of lift-over was wrong.

Methodologically, HDP are highly dependent on the accuracy of SNP identification. HiFi reads instead of ONT reads should be used as HiFiiasm⁵ phases haplotype with HiFi reads, because of the small insertions and base calling error at same bases from ONT and Pacbio CLR reads⁶, lead to false positives in the HDP method, which can be seen in the fluctuating SNP allele number (Supplementary Data 1) instead of the ratio of 1:1 between haplotype. The two regions that are replicated (dup1 vs. dup1r, dup2 vs. dup2r in

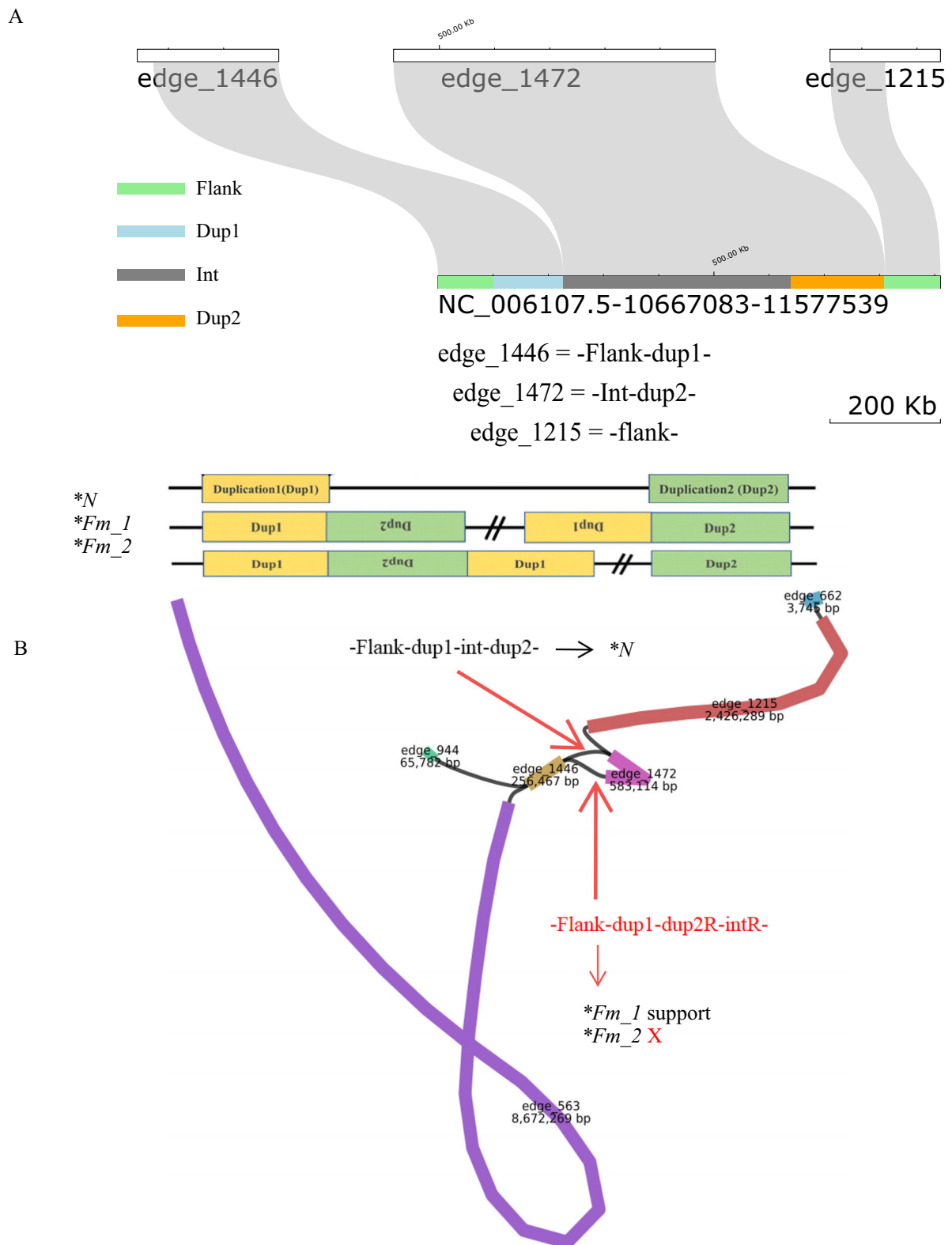


Fig. 1 | The String Graph assembled from Flye supports the scenario **Fm_1* and **N*. **A** The contigs from the graph are related with the *Fm* locus. **B** The String Graph assembled from Flye around edge_1472, which indicates two directions

from edge_1472 connected with edge_1446. e.g. It indicates -flank-dup1-int-dup2- and -Flank-dup1-dup2R-intR, R: reverted, supporting **N* and **Fm_1*, respectively.

Fig 1b¹) have a high degree of similarity to the region that was replicated before (over 99.8%, see Supplementary Data 2, Figure S8). SNP detection in repetitive regions has not been solved too well so far, which adds to the difficulty of the HDP approach. It is also worth noting that Nagarjun

mistakenly referred to HiFi raw reads (~39X)¹ as PacBio long reads (they used >600X as subreads instead of HiFi CCS reads) based on the description in the Method⁴, where in fact Nagarjun only used ONT long reads in their work.

To validate of the reliability of the *Fm* locus in GRCg6a that they used for phasing, it is not sufficient to rely solely on tiling paths. A gold standard for assessing the accuracy of the assembly structure in the target region is to verify the reads by aligning them back to the reference genome and checking that the depth of this region is similar to the depth across the entire genome^{7,8}. Based on our work, there is a zero-depth window of breakpoint (Figure S9–10) in the GRCg6a *Fm* locus with Ultra-Long reads (UL) and HiFi reads from wild-type non-black-bone HuXu breeds (PRJNA693184), indicating an inappropriate structure of the *Fm* locus in GRCg6a.

Finally, it seems that they didn't take into account that genome assembly includes both pre-assembly read correction and post-assembly polishing steps (Figure S1 in our work¹) when they doubted several base pairs did not match the reference sequence. During polishing, contigs/scaffolds were first polished by ONT reads and then by HiFi reads. For certain bases, there may be inconsistencies between the ONT reads and the final assembly.

Method

The region used for mapping as reference was defined as 100 kb upstream and downstream of the *Fm* locus (CAU_Silkie, CM065509.1:10916607-12125334; GRCg6a, NC_006107.5:10667083-11577539). Minimap2 (2.26-r1175)⁹ and Samtools (1.15.1)¹⁰ were used for reads mapping. For depth calculation: `minimap2 --secondary=no --sam-hit-only --MD -Y -t $threads -ax map-HiFi/map-ont $ref $reads | samtools view -e "rlen/qlen >=0.99" -F 256 -b -h -@ $threads - | samtools sort -m 2 G -@ $threads -o $o -`. For heterozygote validation: `minimap2 --secondary=no --sam-hit-only --MD -Y -t $threads -ax map-ont $ref $reads | samtools view -e "rlen/qlen >=0.99 && rlen>50000" -F 256 -b -h -@ $threads - | samtools sort -m 2 G -@ $threads -o $o -`. Online version of LINKVIEW2 was used for visualization of reads and contigs alignments. The visualization of the String Graph is done using Bandage (0.8.1)¹¹, nodes around edge_1472 was used for visualization. Clipped reads visualization was accomplished by IGV (2.16.2)³. The depth calculation is finished with mosdepth (0.3.3)¹² by `'-n -b 100'`.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All genome assembly datasets reported in this study have been deposited in GenBank (NCBI) under accession numbers PRJNA805080 and PRJNA827662. This WGS project of Silkie chicken has been deposited at DDBJ/ENA/GenBank under the accession JAKZEP000000000. The version described in this paper is version JAKZEP010000000. The data of GGswu (huxu breed) is available under the NCBI accession PRJNA693184.

Received: 27 March 2024; Accepted: 26 February 2025;

Published online: 01 April 2025

References

- Zhu, F. et al. A chromosome-level genome assembly for the Silkie chicken resolves complete sequences for key chicken metabolic, reproductive, and immunity genes. *Commun. Biol.* **6**, (2023).
- Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
- Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant Review with the Integrative Genomics Viewer. *Cancer Res.* **77**, E31–E34 (2017).
- Shinde, S. S., Sharma A., & Vijay, N. Decoding the fibromelanosis locus complex chromosomal rearrangement of black-bone chicken: genetic differentiation, selective sweeps and protein-coding changes in Kadaknath chicken. *Front. Genet.* **14**, (2023).
- Cheng, H. Y., Concepcion, G. T., Feng, X. W., Zhang, H. W. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with HiFiasm. *Nat. Methods* **18**, 170 (2021).

- Wang, Y. H., Zhao, Y., Bollas, A., Wang, Y. R. & Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* **39**, 1348–1365 (2021).
- Chen, J. et al. A complete telomere-to-telomere assembly of the maize genome. *Nat. Genet.* **55**, 1221–1231 (2023).
- Bi, G. et al. Near telomere-to-telomere genome of the model plant *Physcomitrium patens*. *Nat. Plants* **10**, 327–343 (2024).
- Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
- Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**, 3350–3352 (2015).
- Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).

Author contributions

Qiang-sen Zhao: Data analysis, Investigation, Visualisation, Validation, Writing - original draft, Writing - review & editing. Feng Zhu: Conceptualization, Writing -original draft, Writing - review & editing. Zhuo-cheng Hou: Conceptualization, Resources, Writing - review & editing, Project administration, Supervision.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-07826-1>.

Correspondence and requests for materials should be addressed to Zhuo-cheng Hou.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: George Inglis and Christina Karlsson Rosenthal.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025