What is the correct genomic structure of the complex chromosomal rearrangement at the *Fm* locus in Silkie chicken?

Ashutosh Sharma, Nagarjun Vijay*

Computational Evolutionary Genomics Lab, Department of Biological Sciences, IISER Bhopal, Bhauri, Madhya Pradesh, India

*Corresponding author: nagarjun@iiserb.ac.in

Abstract

The availability of high-quality chromosome-level genome assemblies of an increasing number of avian species holds significant promise for addressing longstanding questions in bird evolution and biology. In a recent issue of *Communications Biology*, Zhu, F., Yin, ZT., Zhao, QS. et al. (ZYZSJ)[1] presented a chromosome-level assembly for the Silkie chicken using a multi-platform high-coverage dataset to obtain accurate and complete sequences spanning the chicken genome. A key finding from their genomic analysis is the reconstruction of the structure of the complex rearrangement at the *Fm* locus, the primary genetic change underlying the rare and conspicuous dermal hyperpigmentation phenotype generally called Fibromelanosis. However, in contrast to their identification of the **Fm_1* scenario (which the authors refer to as FM2) as the correct arrangement at the *Fm* locus, several previously published studies[2–6] claim that **Fm_2* is the valid scenario. Our re-analysis of ZYZSJ's new genome assembly (CAU_Silkie) demonstrates that **Fm_2* is indeed the correct scenario, and the **Fm_1* scenario favoured by ZYZSJ results from an assembly error caused by mosaic haplotypes generated during the de novo assembly step. We recommend that genome projects perform post-assembly validation and correction to safeguard biological interpretations from the impact of assembly artefacts.

Introduction

ARISING FROM Zhu, F., Yin, ZT., Zhao, QS. et al. Communications Biology
https://doi.org/10.1038/s42003-023-05619-y (2023)

The Fibromelanosis phenotype is found in Silkie and other black-bone chicken breeds, such as Yeonsan Ogye and Kadaknath. Study of this intriguing phenotype for over a century has gradually narrowed the genomic location of the underlying genetic mutation to a complex chromosomal rearrangement (*Fm* locus) on chromosome 20[2–4,7–9]. Black-bone chicken breeds

share a common origin of the complex rearrangement at the *Fm* locus[5]. However, the correct genomic structure of this complex chromosomal rearrangement has been challenging to establish based on short-read sequencing, PCR and genetic analysis of crossing experiments alone[10]. Three possible hypothetical scenarios (**Fm_1*, **Fm_2*, and **Fm_3*) have been proposed (see **Fig. 1**) based on the identified rearrangement junction sequences[3]. Genetic analysis of crosses between individuals with Fibromelanosis (*Fm*) and wild-type (**N*) has favoured the second scenario (**Fm_2*)[3,10]. However, subsequent studies utilising limited long-read sequencing for de novo genome assemblies of Yeonsan Ogye and Silkie chicken have not definitively determined which of the three proposed scenarios is correct[11–13].

ZYZSJ generate high-quality long-read sequencing datasets consisting of >80x (Oxford Nanopore Technology: ONT) and >600x coverage (PacBio) along with >170x Hi-C reads from a single *Fm* homozygote (*Fm/Fm*) individual to assemble a chromosome-level genome for the Silkie chicken. The multi-platform high-coverage dataset generated from a single individual is of excellent quality and ideally suited for conclusively resolving the correct genomic structure of the complex rearrangement at the *Fm* locus. These data are expertly assembled using state-of-the-art tools, resulting in one of the most comprehensive genome assemblies of chickens, which recovers many regions previously absent in earlier assemblies. In ZYZSJ's Figs. 1b and Supplementary Fig. 11-15, they identified **Fm_1* scenario (which the authors call FM2) as the accurate arrangement at the *Fm* locus by relying upon this high-quality, chromosome-level genome assembly.

Results

While ZYZSJ used a de novo assembly approach to discern the correct arrangement at the *Fm* locus, in a parallel study[5], we opted for a haplotype phasing approach using the GRCg6a chicken genome (with a high-quality assembly of the *Fm* locus; see **Supplementary Figures S1-S3**) as an anchor. Despite utilising ZYZSJ's long-read datasets (ONT and PacBio), our analysis consistently supported the **Fm_2* scenario. High-quality chromosome-level assemblies generally serve as invaluable resources, contributing significantly to advances in basic biology research and enhancing our understanding of genomic evolution. Nevertheless, we assert that our haplotype phasing approach excels in resolving the rearrangement at the *Fm* locus for several reasons: (1) we identify specific haplotype-defining positions (HDPs) spanning the entire length

of the *Fm* locus, (2) we quantify long-read support for each pair of HDPs, and (3) we independently validate the results using ONT and PacBio long-reads. Our method's intuitive nature and data visualisation at various steps led us to the correct (**Fm_2*) conclusion, which can be verified by examining the raw data. We observed that the end of the DUP1 region and the start of the DUP2 region have insufficient HDPs (**Fig. 2A**). Therefore, we relied upon haplotype-specific reads that span these HDP-poor regions.

Evaluating long-read support for the CAU_Silkie assembly revealed multiple haplotype switching events (**Fig. 2B**). The absence of raw read support for the published CAU_Silkie genome assembly suggests that the de novo assembly approach has failed to resolve the haplotypes. We observed that the allelic states either did not differ between haplotypes or were inconsistent with the long-read dataset (**Fig. 2B**). Hence, the CAU_Silkie assembly of the *Fm* locus consists of a mosaic haplotype. Due to a mosaic haplotype in CAU_Silkie assembly, the previously[5] identified 49 HDPs were insufficient to perform haplotype phasing. So, we identified 211 new HDPs while re-analysing the long-read data using CAU_Silkie as the reference. All 260 HDPs are concordant with the haplotypes we reconstructed earlier[5], reinforcing our result that the **Fm_2* is the correct structure of the *Fm* locus (**Supplementary Table S1-S6**).

Based on the evidence presented in our re-analysis and the findings from our previously published study[5], we propose that consistent with early studies[3,10], **Fm_2* represents the accurate arrangement at the *Fm* locus, with no support for scenarios **Fm_1* and **Fm_3*. Our evidence that **Fm_2* is the correct structure of the *Fm* locus will be vital to resolving future questions about gene regulation in this region. In a broader context, a black-box approach to genome assembly, lacking downstream validation, can introduce erroneous bases into genome assemblies, potentially impacting evolutionary genomic analysis[14]. Given the increasing capability of long-read sequencing methods to resolve complex structural variations, we advocate for caution when employing genome assembly tools. To enhance the reliability of results, we recommend thorough post-assembly evaluation and correction of the genomic sequence, utilising all raw data used in the assembly process.

Methods

Haplotype-defining positions (HDPs) are those sites that separate haplotype-consistent tiling paths over the entire span of the DUP1 and DUP2 regions and extend into the flanking regions. A

schematic description of HDP identification and haplotype-specific read partitioning using the GRCg6a genome is provided in **Supplementary Figure S4-S5**. Comparing the sequence similarity of the two copies of the duplicated regions assembled in the CAU_Silkie genome with each other and the corresponding region in GRCg6a, we found all three sequences to be highly similar in both DUP1 and DUP2 regions **Supplementary Figure S6-S9**. Hence, using GRCg6a as a reference for haplotype phasing of the Silkie long-read sequences is possible. Nonetheless, we re-analysed the long-read data after mapping it to the CAU_Silkie assembly. In our re-analysis, we scrutinised the CAU_Silkie assembly to assess whether the sequencing raw data substantiated the integrity of the genome assembly. Initially, we examined the allelic states at the 49 HDPs identified in our recent study[5] within the CAU_Silkie assembly after the liftover of positions from GRCg6a. We could not find haplotype-consistent tiling paths based on these HDPs. Considering the possibility that the genome assembler employed a distinct set of HDPs for reconstructing the correct structure of the *Fm* locus, we identified 211 additional sites that differed between the haplotypes of the CAU_Silkie assembly. We adapted our read-backed phasing script to assess long-read support for each pair of these sites. The complete set of HDPs (previous 49 + newly identified 211 positions) consistently recovered the same haplotypes in mapped long reads irrespective of whether the GRCg6a or CAU_Silkie assembly was used.

Acknowledgements

We thank the University Grants Commission for supporting AS with a Ph.D. scholarship. The Department of Biotechnology, Ministry of Science and Technology, India (Grant no. BT/11/IYBA/2018/03) and Science and Engineering Research Board (Grant no. ECR/2017/001430) provided funds for computational resources (i.e., Har Gobind Khorana Computational Biology cluster) used.

Competing interests

The authors declare no competing interests.

Author contributions

Ashutosh Sharma: Conceptualisation, Formal analysis, Investigation, Visualisation, Validation, Writing - original draft, Writing - review & editing. Nagarjun Vijay: Conceptualization,

Resources, Writing - original draft, Writing - review & editing, Funding acquisition, Project administration, Supervision.

Data Availability

Scripts and data are available at https://github.com/Ashu2195/Fm_locus.

References

1. 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. 2023 61* **6**, 1–15 (2023).

2. Dorshorst, B., Okimoto, R. & Ashwell, C. Genomic regions associated with dermal hyperpigmentation, polydactyly and other morphological traits in the silkie chicken. *J. Hered.* (2010).

3. Dorshorst, B. *et al.* A genomic duplication is associated with ectopic eomesodermin expression in the embryonic chicken comb and two duplex-comb phenotypes. *PLoS Genet.* **11**, (2015).

4. Tian, M. *et al.* Genomic regions associated with the sex-linked inhibitor of dermal melanin in Silkie chicken. *Front. Agric. Sci. Eng.* **1**, 242–249 (2014).

5. 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**, 1180658 (2023).

6. Shinomiya, A. *et al.* Gene duplication of endothelin 3 is closely correlated with the hyperpigmentation of the internal organs (fibromelanosis) in silky chickens. *Genetics* (2012).

7. Bateson, B. W., -a, F., Punnett, R. C. & Contents, M. The inheritance of the peculiar pigmentation of the silky fowl. *J. Genet. 1911 13* **1**, 185–203 (1911).

8. Johansson, A. M. & Nelson, R. M. Characterization of genetic diversity and gene mapping in two Swedish local chicken breeds. *Front. Genet.* **6**, 44 (2015).

9. Cha, J. *et al.* Genome-wide association study revealed the genomic regions associated with skin pigmentation in an Ogye x White Leghorn F2 chicken population. *Poult. Sci.* **102**, (2023).

10. Dharmayanthi, A. B. *et al.* The origin and evolution of fibromelanosis in domesticated chickens: Genomic comparison of Indonesian Cemani and Chinese Silkie breeds. *PLoS One* **12**, (2017).

11. Li, M. *et al.* De Novo Assembly of 20 Chicken Genomes Reveals the Undetectable Phenomenon for Thousands of Core Genes on Microchromosomes and Subtelomeric Regions. *Mol. Biol. Evol.* **39**, (2022).

12. Sohn, J. Il *et al.* Whole genome and transcriptome maps of the entirely black native Korean chicken breed Yeonsan Ogye. *Gigascience* **7**, 1–14 (2018).

13. Cho, Y., Kim, J. Y. & Kim, N. Comparative genomics and selection analysis of Yeonsan Ogye black chicken with whole-genome sequencing. *Genomics* **114**, (2022).

14. Mittal, P., Jaiswal, S. K., Vijay, N., Saxena, R. & Sharma, V. K. Comparative analysis of corrected tiger genome provides clues to its neuronal evolution. *Sci. Rep.* **9**, 18459 (2019).

Figure legends

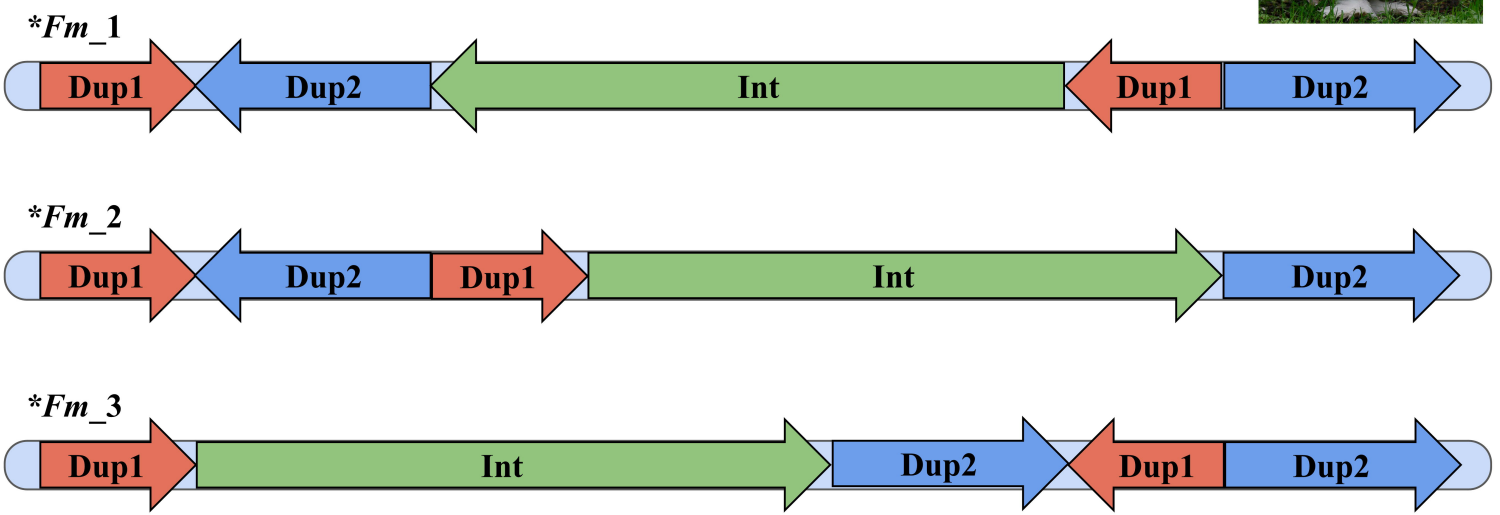
Figure 1: **Organisation of *Fm* locus in Red junglefowl and Silkie black-bone chicken**. The *Fm* locus located on chromosome 20 consists of two distinct non-paralogous regions designated as Duplication 1 (Dup1 ~127 Kb) and Duplication 2 (Dup2 ~170 Kb), separated by an intervening Intermediate (Int ~412 Kb) region. The organisation of the *Fm* locus in non-BBC (**N*) neither involves duplication nor rearrangements. In black-bone chicken breeds, the *Fm* locus has undergone a complex rearrangement event involving duplications and inversions. Although the junctions between these regions have been identified based on short-read sequencing and PCR, the overall organisation of the *Fm* locus has been challenging to decipher. Three possible scenarios (**Fm_1*, **Fm_2*, and **Fm_3*) representing the hypothetical organisation of the *Fm* locus have been proposed by Dorshorst et al., 2011 based on the junction sequences

identified. Dorshorst and most subsequent studies have primarily supported the **Fm_2* scenario. However, some studies have favoured the **Fm_1* scenario.

Figure 2: Inference of haplotypes in Dup1 and Dup2 based on haplotype defining positions (HDP) using ONT and PacBio raw read data. **(A)** Distribution of heterozygous and Haplotype Defining Positions (HDPs) across Dup1 and Dup2. **(B)** The nucleotide bases at each HDP, identified by raw read data phasing, are shown within the rectangular boxes. The haplotypes assembled in the CAU_Silkie genome are displayed beside the rectangular boxes. Notably, HDPs highlighted in green signify collapsed haplotypes (i.e., both haplotypes have the same base in the assembly), while red ones indicate switched haplotypes. The SRA (Short Read Archive) IDs associated with the respective haplotypes cover the parts of Dup1 and Dup2 with a lack of HDPs.

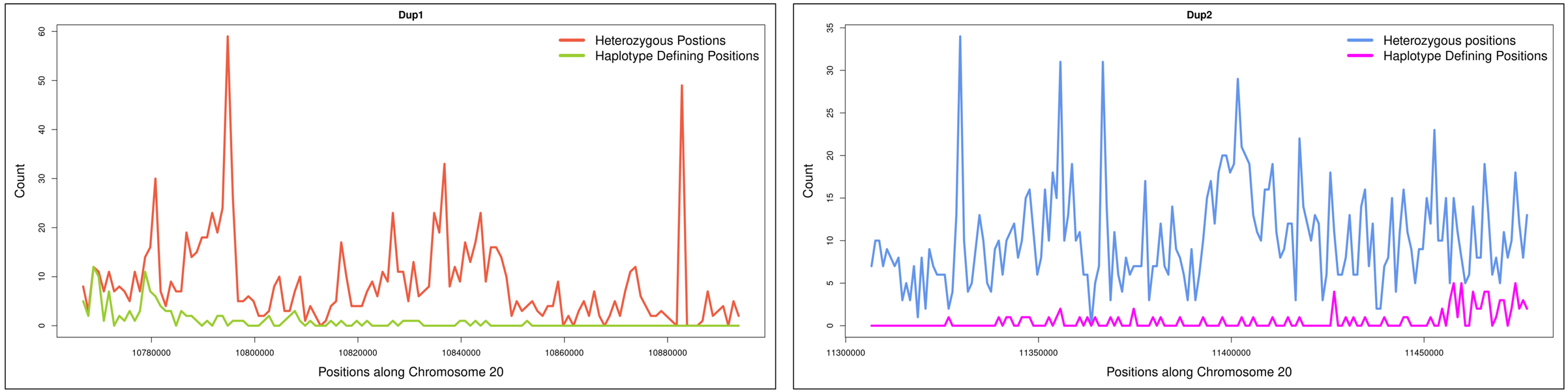
Organisation of *Fm* locus in non-black-bone chickens (*N):

Possible organisations of *Fm* locus in Silkie black-bone chicken:

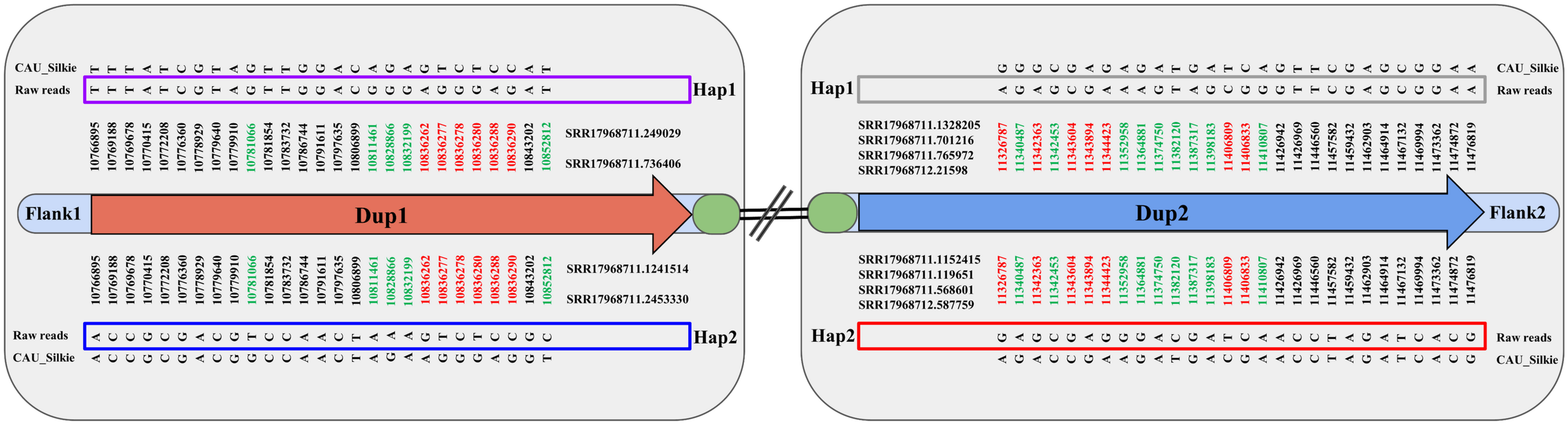


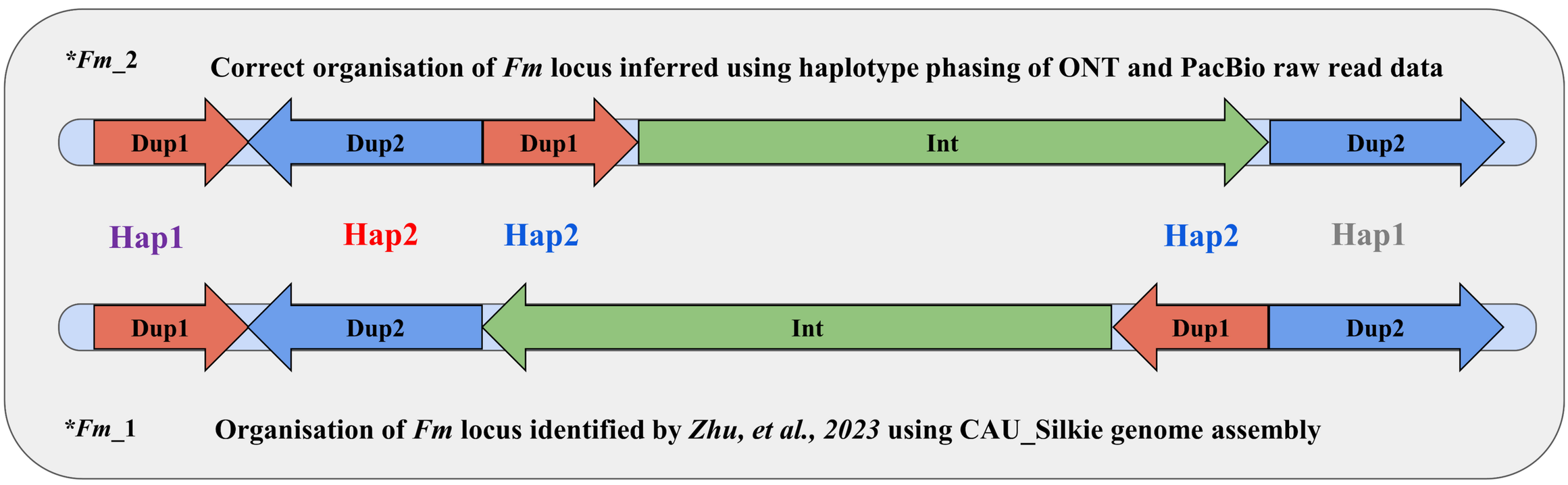
Fm locus Scenarios	Supporting Studies
***Fm*_1**	• Sohn, et al., 2018 • Zhu, et al., 2023
***Fm*_2**	• Dorshorst, et al., 2011 • Dharmayanthi, et al.,2017 • Shinde, et al., 2023
***Fm*_3**	None

A


Dup1
Heterozygous Postions
Haplotype Defining Positions
Count
Positions along Chromosome 20
Dup2
Heterozygous positions
Haplotype Defining Positions
Count
Positions along Chromosome 20

B


CAU_Silkie
Raw reads
Hap1
SRR17968711.249029
SRR17968711.736406
Flank1
Dup1
SRR17968711.1241514
SRR17968711.2453330
Raw reads
CAU_Silkie
Hap2
Hap1
CAU_Silkie
Raw reads
SRR17968711.1328205
SRR17968711.701216
SRR17968711.765972
SRR17968712.21598
Dup2
Flank2
SRR17968711.1152415
SRR17968711.119651
SRR17968711.568601
SRR17968712.587759
Hap2
Raw reads
CAU_Silkie


*Fm_2
Correct organisation of Fm locus inferred using haplotype phasing of ONT and PacBio raw read data
Dup1
Dup2
Dup1
Int
Dup2
Hap1
Hap2
Hap2
Hap2
Hap1
Dup1
Dup2
Int
Dup1
Dup2
*Fm_1
Organisation of Fm locus identified by Zhu, et al., 2023 using CAU_Silkie genome assembly