Characterizing the Major Histocompatibility Complex in the House Finch (*Haemorhous mexicanus*) Using Genomics and Transcriptomics

Jair Torres[1], Tricia Van Laar[1], Joel Slade[2]
[1]Department of Biological Sciences, California State University Stanislaus
[2]Department of Biological Sciences, California State University Fresno

Introduction

The major histocompatibility complex (MHC) is an important gene family responsible for initiating the adaptive immune response in jawed vertebrates[1]. Given the importance of the MHC in immunity, there is evolutionary pressure to diversify this gene family, which can explain why birds exposed to a great variety of pathogens tend to have many MHC gene copies[2]. The most variability in this gene family occurs in MHC class II genes as these encode the beta chain forming the beta chain of the peptide binding groove (Fig. 1). Nevertheless, MHC genes are notoriously difficult to characterize because they are highly variable and subject to duplication and pseudogenization. Long-read sequencing has largely solved this problem[3]. Here, we used a long-read genome assembly of the House Finch (*Haemorhous mexicanus*) to estimate copy number of MHC class II beta genes in these birds and characterize the identified loci.

.

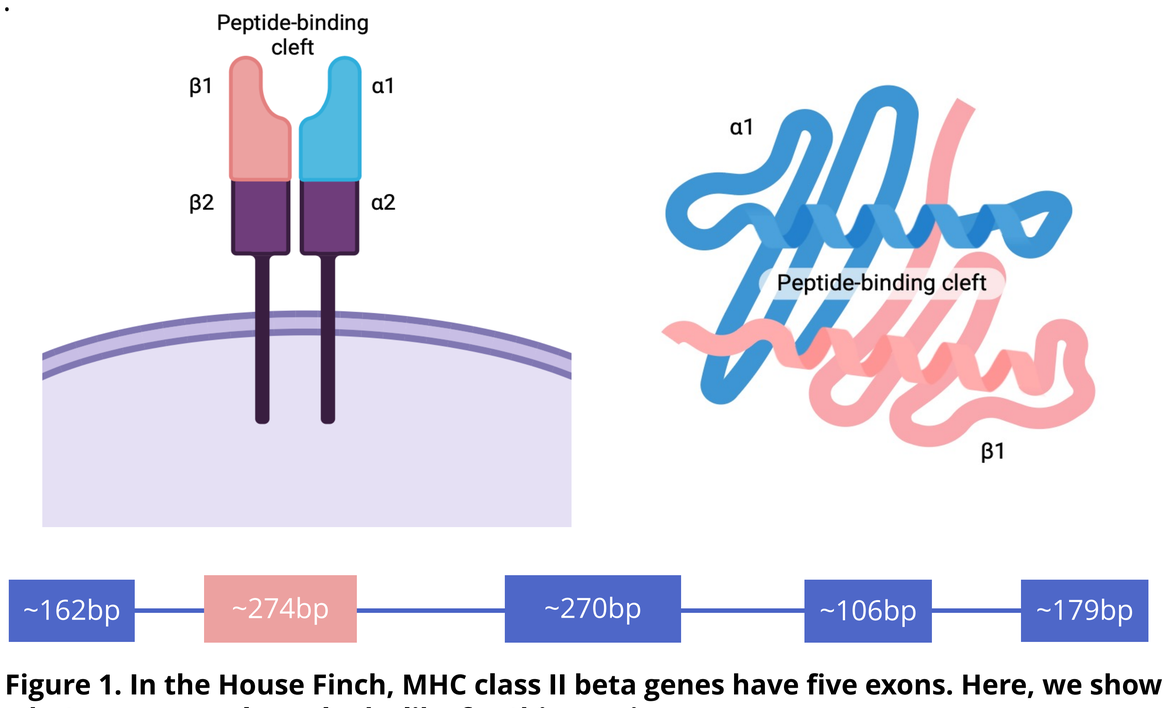

Figure 1. In the House Finch, MHC class II beta genes have five exons. Here, we show what an average locus looks like for this species.

Methods

We assembled the House Finch genome using long-read sequencing data publicly available on NCBI (accession PRJNA1005622) using the genome assembler, hifiasm[4] (Fig. 2). Next, we used BLAST[5] to compare known MHC sequences from the zebra finch (*Taeniopygia guttata*) to identify full-length MHC transcripts in the house finch transcriptome and partitioned these transcripts into exons for alignment against our genome assembly. To annotate the MHCIIB genes, we used the genome visualization software JBrowse[6] to manually curate the alignments by searching for premature stop codons and examining exon boundaries. To verify the exon lengths of our annotations, we used the software SPAdes to construct this bird's transcriptome using RNA-seq data (accession SRP018959).

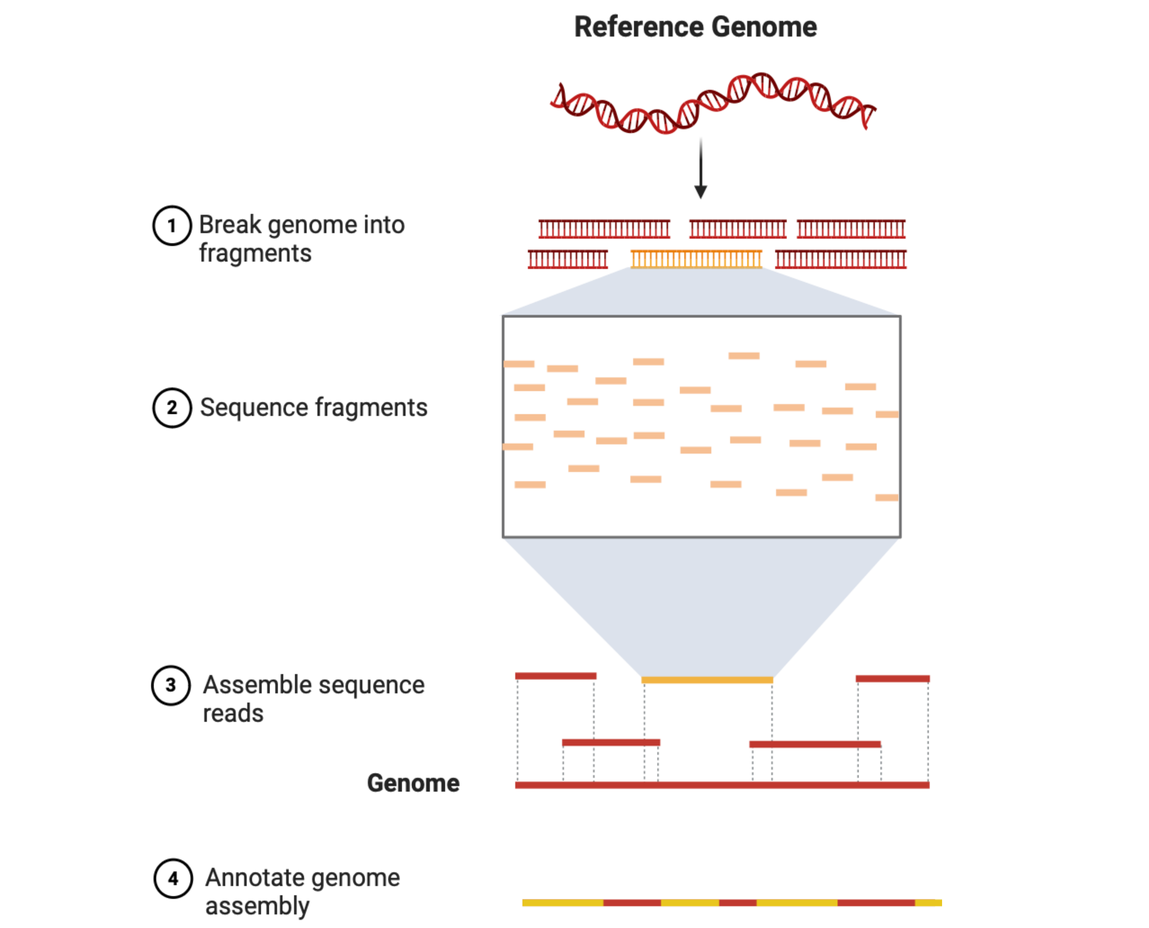

Figure 2. Workflow illustrating the steps used for genome assembly and annotation.

Results

We identified 15 loci across three scaffolds, of which 8 were identified as putatively functional and 7 were classified as pseudogenes due to premature stop codons or missing splice sites. Our findings are consistent with previous studies demonstrating high copy numbers and many pseudogenes in passerine birds. These findings also show copy number variation among individuals of the same species as one parental haplotype of this bird had 15 copies of the MHCIIB gene while the other had 12.

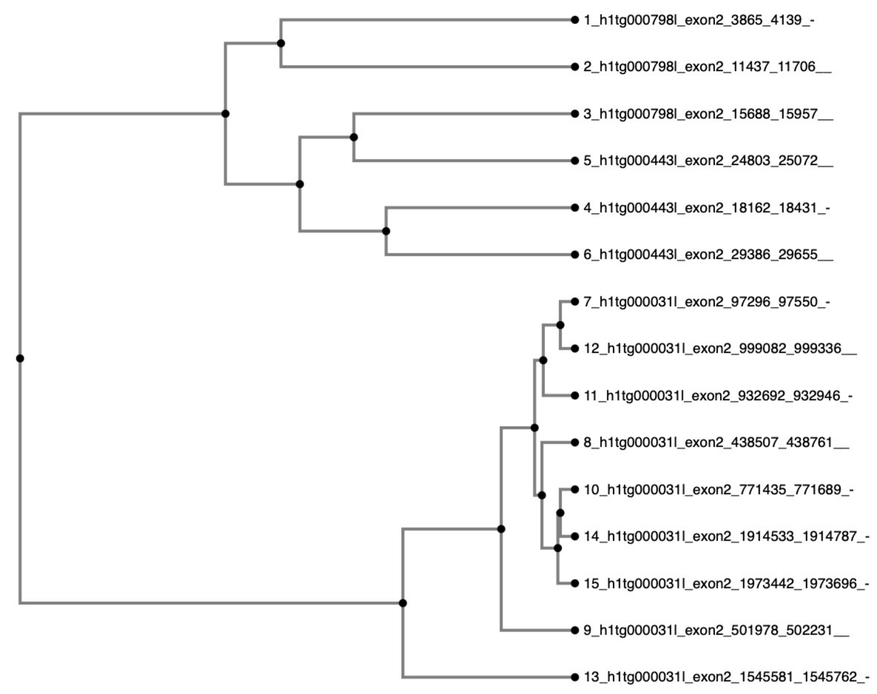

Figure 3. UPGMA dendrogram of the MHCIIB genes from one parental haplotype.

Discussion

Our findings are consistent with previous studies showing high copy numbers and many pseudogenes in passerine birds. Interestingly, the House Finch had a high MHC copy number despite being non-migratory. Comparative studies between House Finch populations in its native and introduced areas would provide valuable insights into how differing pathogen exposure and evolutionary pressures influence MHC class II beta diversity. Our study contributes to our understanding of MHC diversity in non-model species and demonstrates the strengths of long-read sequencing in accurately characterizing complex genomic regions which can inform future immunogenetic research.

References

1. Kaufman, J. (2018). Unfinished business: evolution of the MHC and the adaptive immune system of jawed vertebrates. *Annual review of immunology*, *36*(1), 383-409.
2. Borghans, J. A., Beltman, J. B., & De Boer, R. J. (2004). MHC polymorphism under host-pathogen coevolution. *Immunogenetics*, *55*, 732-739.
3. He, K., Minias, P., & Dunn, P. O. (2021). Long-read genome assemblies reveal extraordinary variation in the number and structure of MHC loci in birds. *Genome Biology and Evolution*, *13*(2).
4. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature methods*, *18*(2), 170-175.
5. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410.
6. Buels, R., Yao, E., Diesh, C. M., Hayes, R. D., Munoz-Torres, M., Helt, G., ... & Holmes, I. H. (2016). JBrowse: a dynamic web platform for genome visualization and analysis. *Genome biology*, *17*, 1-12.
7. Bushmanova, E., Antipov, D., Lapidus, A., & Prjibelski, A. D. (2019). rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *GigaScience*, *8*(9), giz100.

Acknowledgements

This project was supported by CSU Stanislaus RSCA program and the CSUBIOTECH Faculty-Graduate Student Research Collaboration Grant. Illustrations in this poster were created with BioRender (biorender.com).

CALIFORNIA STATE UNIVERSITY STANISLAUS

Visit Our GitHub