

Project 6: Enabling comparative genomics of heterogametic sex chromosomes

Kamil Jaron, Sanger (Lead)

Emily Clark, EMBL-EBI (Co-Supervisor)

Sex chromosomes are of a major interest to evolutionary biologists as they challenge our general understanding of the effects of ploidy and recombination¹. In many clades, recombination between the two sex chromosomes is arrested resulting in a large non-recombining region on heterogametic sex chromosomes (Y or W) that gradually degenerates. Understanding this dynamic was mostly restricted to relatively recent sex chromosome changes, where the degeneration occurs in the early stages and the heterogametic sex chromosome is not yet too repetitive. Large comparative genomic studies of sex chromosomes are vastly restricted to homogametic sex chromosomes, due to challenges with assembling repetitive heterogametic chromosomes.

Sex chromosome sequence is typically derived from a genome assembly of the heterogametic sex and post-processing of the scaffolds using reads from one or both sexes^{2,3}. Current assembly techniques assume a constant ploidy (usually diploidy) across the whole genome and struggle with divergent and frequently repetitive sex chromosomes^{4,5}. In this project, we will develop an approach that integrates both male and female sequencing data directly into an assembly graph, enabling a more direct way of sex chromosome assembly. Integration will be facilitated through *k*-mer comparison of the male and female sequencing libraries to generate sets of sex-chromosome markers⁶ that will be used for marking up and interpretation of the assembly graph. Hierarchical interpretation and subtraction of paths on chromosome-basis has the ambition of greatly lowering the complexity of the assembly graph and might dramatically improve continuity of assembled sequences. Here we aim to develop a proof-of-concept assembler that might be transformative in our ability to analyse sex chromosomes among diverse systems.

Nearly all Diptera feature male heterogamety⁷ (XY sex chromosomes), therefore we can complete the full genome using a single male individual of a species. Diptera, feature conservation of the canonical X chromosome, as well as numerous changes in different dipteran lineages^{2,8,9}. Y chromosomes, on the other hand, have only very little known about them. Distantly related species have no detectable homology on gene level, although they do show signatures of convergent evolution¹⁰. To enable a more reliable representation of X and Y chromosomes, Kamil Jaron's group is sequencing male and female specimens for >100 species. The richness of the dataset will provide a perfect opportunity for testing the newly developed approach to understand the heterogametic sex dynamic across the order. Furthermore, Y chromosomes are expected to be enriched in male-fuction, which are frequently young and fast evolving genes that are challenging to annotate and analyse and not captured by orthologous gene set analysis such as BUSCO. Emily Clark's research group and the Ensembl Compara team would support the project through annotation of novel genes and Y-linked genes on the Y-chromosomes, and in performing presence/absence analysis of these genes across Diptera species. Ensembl Compara have a deployable version of the FastOMA pipeline¹¹, which uses a *k*-mer based approach to infer orthologous gene relationships, which could be applied for this project. Optimal alignment approaches will also be explored through collaboration between the Clark and Jaron groups using existing and new workflows to investigate intactness of the Y-chromosome across Diptera species. Comparative analysis of novel and Y-linked genes on the Y-chromosomes will contribute to understanding whether they are restricted to certain clades, or unique to a species, enabling applications in phylogenetics, evolutionary biology and biodiversity monitoring.

This project is an **intersection of biologically informed method development and large-scale genomics analysis** enabling both testing of the approach and enabling it to address biological questions such as Y chromosome dynamics during sex chromosome changes, speed of gene degeneration.

The applicant would perform the initial research of the method in Jaron's team. Once a prototype is made, the candidate would have the opportunity to work closely with the Compara team within the Ensembl project and gain experience in service provision, and how service teams operate to deliver community genomics resources. As such the candidate would be well equipped to further advance their research career in any of the listed directions. Soft skills developed during the project would include management of large and complex projects and cross team collaboration towards project delivery. In addition, there would be the opportunity for the applicant to be involved in the outreach components of the Ensembl project developing and providing training for comparative genomics resources, which could utilise the project outcomes as an example. This would contribute to the development of an interdisciplinary portfolio including bioinformatic method development, comparative genomics, reproductive and evolutionary biology.

References

1. Saunders, P. A. & Muyle, A. Sex Chromosome Evolution: Hallmarks and Question Marks. *Mol. Biol. Evol.* **41**, msae218 (2024).
2. Vicoso, B. & Bachtrog, D. Numerous Transitions of Sex Chromosomes in Diptera. *PLOS Biol.* **13**, e1002078 (2015).
3. Palmer, D. H., Rogers, T. F., Dean, R. & Wright, A. E. How to identify sex chromosomes and their turnover. *Mol. Ecol.* **28**, 4709–4724 (2019).
4. Rhie, A. *et al.* The complete sequence of a human Y chromosome. *Nature* **621**, 344–354 (2023).
5. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
6. Jenike, K. M. *et al.* k-mer approaches for biodiversity genomics. *Genome Res.* <https://doi.org/10.1101/gr.279452.124> (2025) doi:10.1101/gr.279452.124.
7. Morelli, M. W., Blackmon, H. & Hjelman, C. E. Diptera and Drosophila Karyotype Databases: A Useful Dataset to Guide Evolutionary and Genomic Studies. *Front. Ecol. Evol.* **10**, (2022).
8. Layana, L., Toups, M. A. & Vicoso, B. Causes and consequences of sex-chromosome turnovers in Diptera. *Evol. Lett.* qrag003 (2026) doi:10.1093/evlett/qrag003.
9. Gries, J. *et al.* 340 dipteran genomes reveal the origin of Muller elements and sex chromosomes in Diptera. 2026.06.01.729285 Preprint at <https://doi.org/10.64898/2026.06.01.729285> (2026).
10. Mahajan, S. & Bachtrog, D. Convergent evolution of Y chromosome gene content in flies. *Nat. Commun.* **8**, 785 (2017).
11. Majidian, S., Nevers, Y., Yazdizadeh Kharrazi, A. *et al.* Orthology inference at scale with FastOMA. *Nat Methods* **22**, 269–272 (2025). <https://doi.org/10.1038/s41592-024-02552-8>