**Annotated Bibliography CG2**

**Amanda Markee**

Baker, R. H., Corvelo, A., & Hayashi, C. Y. (2022). Rapid molecular diversification and homogenization of clustered major ampullate silk genes in Argiope garden spiders. *PLOS Genetics*, *18*(12), e1010537. <https://doi.org/10.1371/journal.pgen.1010537>

* Annotation: Paper that looks into the gene structure and gene duplication in long, repetitive silk genes. This paper uses long-reads from Oxford Nanopore, and annotates the genes for building a tree of sequence relationships.

Ellis, E. A., Storer, C. G., & Kawahara, A. Y. (2021). *De novo* genome assemblies of butterflies. *GigaScience*, *10*(6), giab041. <https://doi.org/10.1093/gigascience/giab041>

* Annotation: This paper reviews existing next-generation genomes available for Lepidoptera as of 2021. A majority of these sequences were short read, and the authors assemble short reads from the SRA database. They measure quality for short read assemblies and assess contributing factors.

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9451–9457. <https://doi.org/10.1073/pnas.1921046117>

* Annotation: RepeatModeler is a software used in whole genome annotation, specifically for identifying repeat elements that should be masked prior to genome analysis. Many new long-read whole genome references use this program to model and mask repeats, especially in Darwin Tree of Life.

Godfrey, R. K., Britton, S. E., Mishra, S., Goldberg, J. K., & Kawahara, A. Y. (2023). A high-quality, long-read genome assembly of the whitelined sphinx moth (Lepidoptera: Sphingidae: Hyles lineata) shows highly conserved melanin synthesis pathway genes. *G3 Genes|Genomes|Genetics*, *13*(6), jkad090. <https://doi.org/10.1093/g3journal/jkad090>

* Annotation: A new reference genome for the hawkmoth *Hyles lineata*. This paper produces a long-read reference genome, and briefly analyzes the differential gene expression in olfaction genes between sexes.

Heckenhauer, J., Stewart, R. J., Ríos-Touma, B., Powell, A., Dorji, T., Frandsen, P. B., & Pauls, S. U. (2023). Characterization of the primary structure of the major silk gene, h-fibroin, across caddisfly (Trichoptera) suborders. *iScience*, *26*(8), 107253. <https://doi.org/10.1016/j.isci.2023.107253>

* Annotation: This study characterizes the gene structure of the primary silk gene in caddisflies. A novel study in describing the genomic architecture of a repetitively challenging gene.

Hotaling, S., Sproul, J. S., Heckenhauer, J., Powell, A., Larracuente, A. M., Pauls, S. U., Kelley, J. L., & Frandsen, P. B. (2021). Long Reads Are Revolutionizing 20 Years of Insect Genome Sequencing. *Genome Biology and Evolution*, *13*(8), evab138. <https://doi.org/10.1093/gbe/evab138>

* Annotation: This review paper discusses the changes in the field of insect genomics over the past 20 years with the introduction of long-read sequencing, in terms of genome production, quality, and other metrics.

Kawahara, A. Y., Storer, C. G., Markee, A., Heckenhauer, J., Powell, A., Plotkin, D., Hotaling, S., Cleland, T. P., Dikow, R. B., Dikow, T., Kuranishi, R. B., Messcher, R., Pauls, S. U., Stewart, R. J., Tojo, K., Frandsen, P. B., Storer, C. G., Markee, A., Heckenhauer, J., … Frandsen, P. B. (2022). Long-read HiFi sequencing correctly assembles repetitive heavy fibroin silk genes in new moth and caddisfly genomes. *Gigabyte*, *2022*, 1–14. <https://doi.org/10.46471/gigabyte.64>

* Annotation: First long-read genome assembly and annotation of silk genes for a pantry moth, and first to compare two orders of insects. This prompted several other comparative silk analyses using long-read sequencing.

*Long-read human genome sequencing and its applications | Nature Reviews Genetics*. (n.d.). Retrieved November 14, 2024, from <https://www.nature.com/articles/s41576-020-0236-x>

* Annotation: This relatively new nature review discusses long-read sequencing in the world of human genomics, and some of the novel research questions and approaches being answered involving disease research, isoform discovery, and gap filling.

Markee, A., Godfrey, R. K., Frandsen, P. B., Weng, Y.-M., Triant, D. A., & Kawahara, A. Y. (2024). De Novo Long-Read Genome Assembly and Annotation of the Luna Moth ( *Actias luna* ) Fully Resolves Repeat-Rich Silk Genes. *Genome Biology and Evolution*, *16*(7), evae148. <https://doi.org/10.1093/gbe/evae148>

* Annotation: Example paper for how long-read sequencing has been used in genomics. A new genome assembly and annotation for the luna moth which compares both short read NGS sequencing, and long read TGS. Long read sequencing produces higher quality and completeness.

Marx, V. (2023). Method of the year: Long-read sequencing. *Nature Methods*, *20*(1), 6–11. <https://doi.org/10.1038/s41592-022-01730-w>

* Annotation: Nature review discussing the expansion in research due to long-read sequencing. Deems this method as method of the year, in the sense that now several interdisciplinary studies have produced publication using this.

Powell, A., Heckenhauer, J., Pauls, S. U., Ríos-Touma, B., Kuranishi, R. B., Holzenthal, R. W., Razuri-Gonzales, E., Bybee, S., & Frandsen, P. B. (2024). Evolution of Opsin Genes in Caddisflies (Insecta: Trichoptera). *Genome Biology and Evolution*, *16*(9), evae185. <https://doi.org/10.1093/gbe/evae185>

* Annotation: A new phylogenetic tree of newly discovered opsins and opsin paralogs in caddisfly groups. A new application for long-read sequencing. This paper also measures selection on lineages with gene duplication.