Chromosome-Scale Assembly Reveals Complex Chromosomal Fusions in a Fritillary Genome

Chromosomal fusion events in a species of butterfly were revealed using Arima Hi-C technology and whole genome sequencing.

Genome assembly is a critical foundational tool for scientific research in biodiversity, conservation, and evolutionary biology. Complete, reference-quality genomes can act as points of reference for the structure and organization of a species' genome and aid in understanding speciation, population-wide genetic diversity, and conservation efforts[1]. These high-quality genomes add immense biological value and are now being regularly generated through larger international consortia including the Vertebrate Genome Project and the Darwin Tree of Life[2].

Unfortunately, whole genome sequencing (WGS) alone, while critical for generating accurate sequence information, does not produce a complete, chromosome-scale assembly. These chromosome-scale assemblies can only be achieved with the addition of a scaffolding technology, such as Arima Hi-C, which enables the ordering and orienting of contigs and anchoring those contigs to chromosomes.

Here, we present the work of a group of scientists who used a chromosome-scale assembly to provide new biological insights into the evolution of butterflies.

Challenge: Characterizing chromosomal evolution in *Brenthis ino*

Scientists have developed an approach that couples WGS and Arima Hi-C technologies to produce a robust chromosome-level genome assembly for the lesser marbled fritillary, *Brenthis ino* **(Figure 1)**. Although most butterflies in the Nymphalidae family have 31 pairs of chromosomes, *B. ino* only has 12-14 chromosomes. Understanding how chromosomes have evolved in this species requires further investigation due to the absence of a genome assembly for these species.

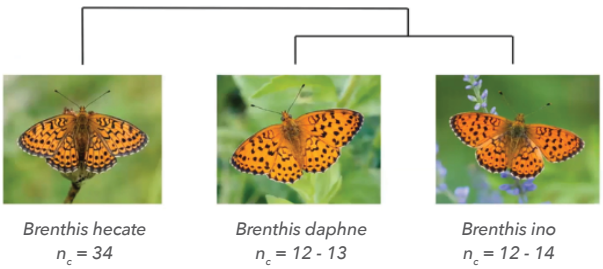

Figure 1. Phylogenetic relationship between three *Brenthis* species. Typically, Lepidoptera have 31 chromosomes, whereas *B. daphne and B. ino* have only 12-14 chromosomes[3]. Photographs are by Vlad Dincă.

To meet this challenge, a research team based out of the University of Edinburgh harnessed the power of Arima Hi-C technology and WGS to produce a complete genome assembly for *B. ino*. By combining 3D genomics and WGS, the team uncovered important mechanisms of genome evolution.

Technology

The Arima Genome Assembly HiC kit enables generation of chromosome-scale, reference-quality assemblies using a range of sample types.

Approach: Profiling the genome sequence and structure of *Brenthis ino*

The team began by producing DNA sequencing libraries from two male *B. ino* individuals and one female *B. ino* individual collected in Somiedo, Braña de Mumian, Asturias, Spain **(Figure 2)**. One of the male individuals was used to generate PacBio continue long reads (CLR) and Illumina TruSeq paired-end read data. The *B. ino* genome was first assembled using the CLR reads. Then, consensus sequencing errors from the resulting contigs were corrected by aligning the Illumina WGS reads against these contigs. This approach resulted in an initial genome assembly of 119 contigs with a total length of 411.8 Mb.

Figure 2. Fore and hind wings of the two male *B. ino* butterflies used to generate the *B. ino* genome sequences[3]. A-B represent the male butterfly used to generate PacBio and Illumina WGS reads, while C-D represent the male butterfly used to generate Hi-C reads.

To ascertain the 3D genomic structure of the *B. ino* genome, researchers next leveraged the Arima Hi-C kit on the second *B. ino* male. To identify interactions within and between the male *B. ino* chromosomes, the trimmed Hi-C reads were aligned against the assembled contigs generated by WGS. The contigs were then scaffolded into 14 chromosome-level sequences. Synteny analyses between the *B. ino* and *M. cinxia* genomes revealed many chromosomal fusions, as well as fissions or reciprocal translocations, that have shaped the *B. ino* genome **(Figure 3)**.

Notably, the Hi-C reads revealed enrichment of genomic contacts between chromosomes 11 and 13 in the *B. ino* genome. Normalized read coverage plots of chromosomes 11 and 13 between male and female genomes revealed that chromosome 11 was the sex chromosome, while chromosome 13 was an autosomal chromosome **(Figure 4)**.

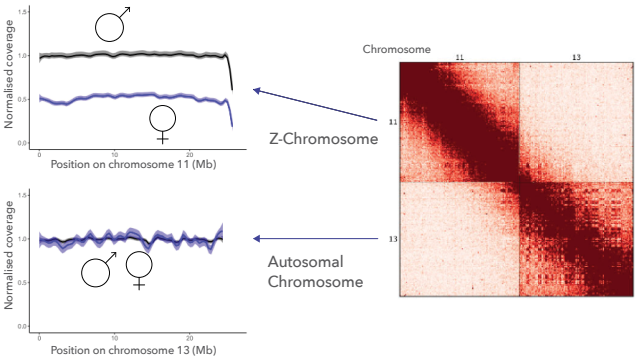

Figure 4. An interchromosomal interaction between the Z-chromosome and an autosomal chromosome. Left: Normalized coverage of WGS reads from chromosomes 11 and 13 of the male (black) and female (blue). Right: Hi-C contact heatmaps for the *B. ino* genome assembly at chromosomes 11 and 13.

Despite observing Hi-C contacts between the Z chromosome and chromosome 13, the strength of the signal remained below what was typically observed within chromosomes in the dataset. As such, whether the enrichment of these contacts represented a fusion between these chromosomes remained unclear. To answer this question, the researchers generated haplotype-specific Hi-C maps by parsing the reads containing SNPs present in either haplotype, which enabled them to observe Hi-C contacts specific to each haplotype **(Figure 5)**. Taken together, the researchers determined that the male individual from which the Hi-C library was generated was heterozygous for a Z-autosome chromosome fusion, i.e. a neo-Z chromosome.

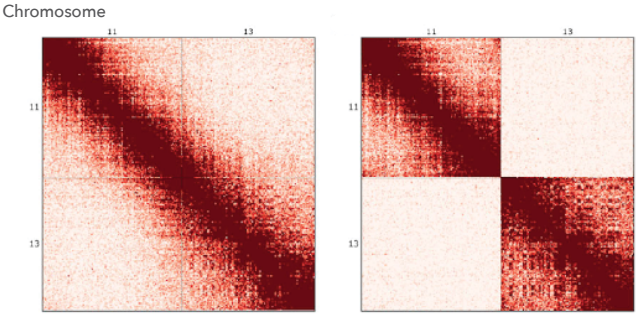

Figure 5. Contact heat maps for Hi-C reads containing alleles exclusively associated with haplotype 1 (left) and haplotype 2 (right).

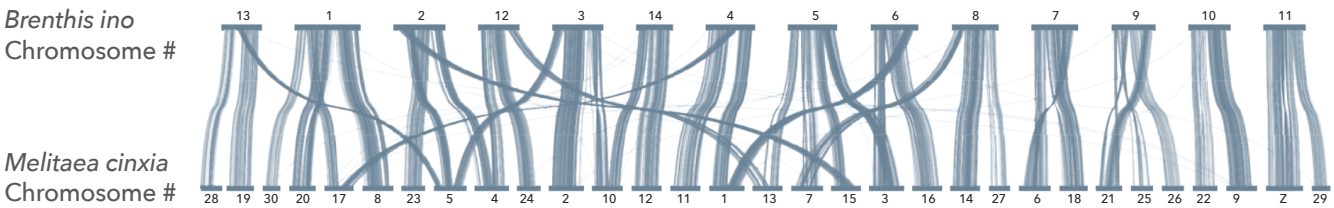

Figure 3. Synteny comparison between *B. ino* and *M. cinxia* chromosomes. Each line connects a BUSCO gene observed in either genome assembly.

Impact: Insights into chromosome evolution in the lesser marbled fritillary

In this study, scientists investigated chromosome evolution in *B. ino* by generating a reference-quality genome assembly. The assembly provided evidence of extensive rearrangements within the *B. ino* genome and haplotype information in the Arima Hi-C data allowed the detection of a heterozygous neo-Z chromosome.

The Arima Hi-C kit augments existing WGS pipelines by providing a picture of 3D genome organization. Utilization of the Arima Hi-C kit allowed the authors to shed light on mechanisms of chromosome evolution in the genus Brenthis. Furthermore, 3D genome mapping can be applied to other eukaryotic organisms to better understand evolutionary mechanisms across the animal kingdom. Finally, with over a million plant and animal species at risk of extinction, efforts to understand 3D genomics will enhance conservation initiatives and ensure environmental sustainability.

> "The Arima Hi-C kit provides a powerful approach for studying genomics as the data contains information about 3D genome organization as well as differences between haplotypes. For example, we identified a heterozygous chromosome rearrangement within a single individual using Hi-C. The Arima HiC data was crucial for us to gain an understanding of chromosome evolution in our study organisms, the lesser marbled fritillary, *Brenthis ino*."
>
> – Alexander Mackintosh, PhD Student
> University of Edinburgh

References

1. Brandies, P. et al., (2019). The value of reference genomes in the conservation of threatened species. *Genes*, 10(11), 846.
2. Rhie, A., et al. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 592, 737–746.
3. Mackintosh, A., et al., (2022). The genome sequence of the lesser marbled fritillary, *Brenthis ino* , and evidence for a segregating neo-Z chromosome. *G3 Genes|Genomes|Genetics*, 12(6), jkac069.

Learn More Webinar: Where Will Genomes Take Us Next: How Chromosome-Scale Assemblies Are Unlocking New Biology

arimagenomics.com | team@arimagenomics.com

CS-003-110922

ARIMA GENOMICS