WAA genome assembly QC summary

Our *de novo* genome assembly of woolly apple aphid (WAA, *Eriosoma lanigerum*) was compared against the WAA genome assembly published by Biello et al. (2020). The two genomes are both assembled using a combination of 10X Genomics linked-reads and *in vivo* Hi-C data. Our final genome assembly is 327 Mb, which is slightly smaller than the 335 Mb genome from Biello et al. (2020). The contig and scaffold N50 values of our assembly are 784 kb and 6 Mb, respectively, which is more fragmented compared to Biello et al. (2020, scaffold N50 71 Mb). The assembly include 92% of the 303 conserved eukaryote single-copy orthologues based on BUSCO analysis, which is also lower than the 97% BUSCO completeness of the other assembly. Finally, there are currently 14,769 predicted protein-coding genes from the manufacturer’s annotation pipeline, which is also only 52% of the 28,186 genes published by Biello et al. (2020). In summary, the genome assembly of Biello et al. (2020) is more complete and more continuous than ours and their annotation also has higher quality.

On the other hand, genome alignment has demonstrated interesting patterns of large-scale inversions between the two assemblies (Fig. 1). These inversions are largely located on the ends of scaffolds, suggesting the possibility of assembly error. Further validation of these inversions may require additional long-read sequencing or bioinformatic analysis using the linked reads from 10X from both assemblies. In addition, sequence identities are unevenly distributed across scaffolds, suggesting that some chromosomes contain more SNPs than others.

Chart, scatter chart

Description automatically generated

Fig. 1 Whole genome alignment of the two assemblies of WAA. Alignment conducted by minimap2 and visualized by dotplotly.