

We downloaded a variant callings data of the *Arabidopsis thaliana* Bur_0 accession from <http://mtweb.cs.ucl.ac.uk/mus/www/19genomes/variants.SDI>. We generated a query genome and a benchmark alignment by replacing the TAIR10 reference alleles with alternative alleles. To systematically understand the feature of the AnchorWave program in detail, we compared the alignment result of AnchorWave with the benchmark alignment and the alignments from other alignment programs. By viewing the alignments manually via IGV (<https://software.broadinstitute.org/software/igv/>) from the first base-pair of Chr1, we checked 50 variant regions where alignment programs generated different alignments. Some of the inconsistent alignment from alternative alignment (for what is alternative alignment, please refer to <https://dx.doi.org/10.1093%2Fbioinformatics%2Fbtv112>, <https://dx.doi.org/10.3389%2Ffgene.2019.01046>, <https://doi.org/10.1371/journal.pgen.1007699>) were skipped, due to they are so common.

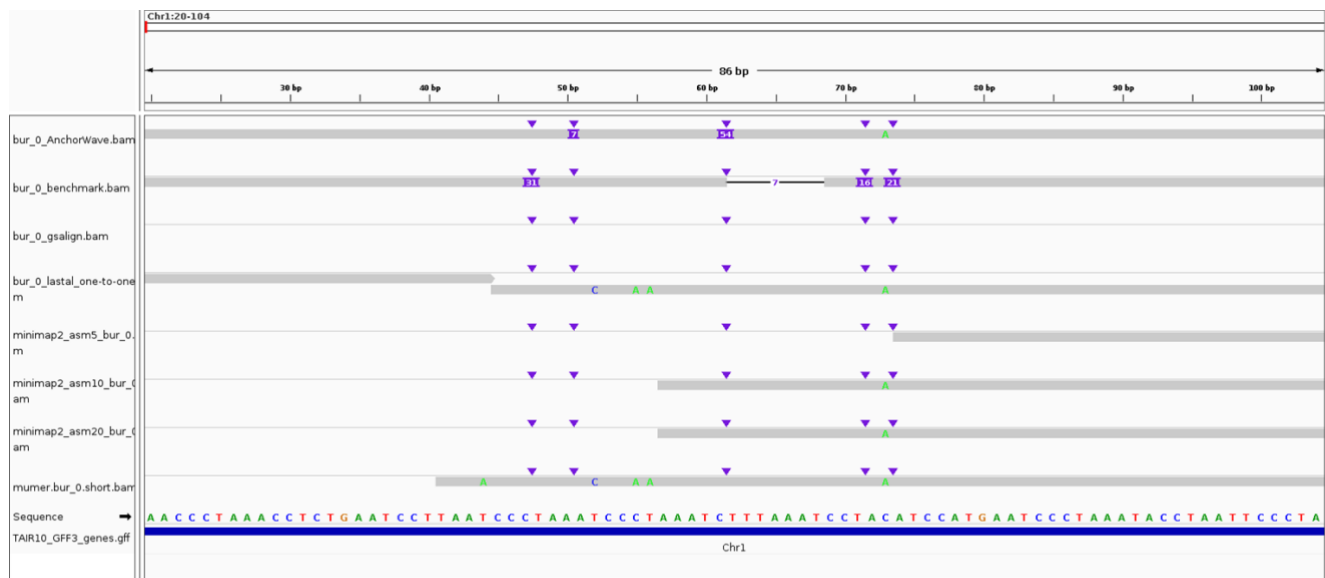


Fig. E1. In this window, AnchorWave identified two INDELs and an SNP. The benchmark reported 4 INDELs. While other programs did not cover the reference genome or query genome completely, likely due to the usage of local alignment strategy. In the perspective of sequence alignment scoring and the principle of parsimony, the alignment of AnchorWave is more optimized than the benchmark alignment (two indels and an SNP v.s. four indels).

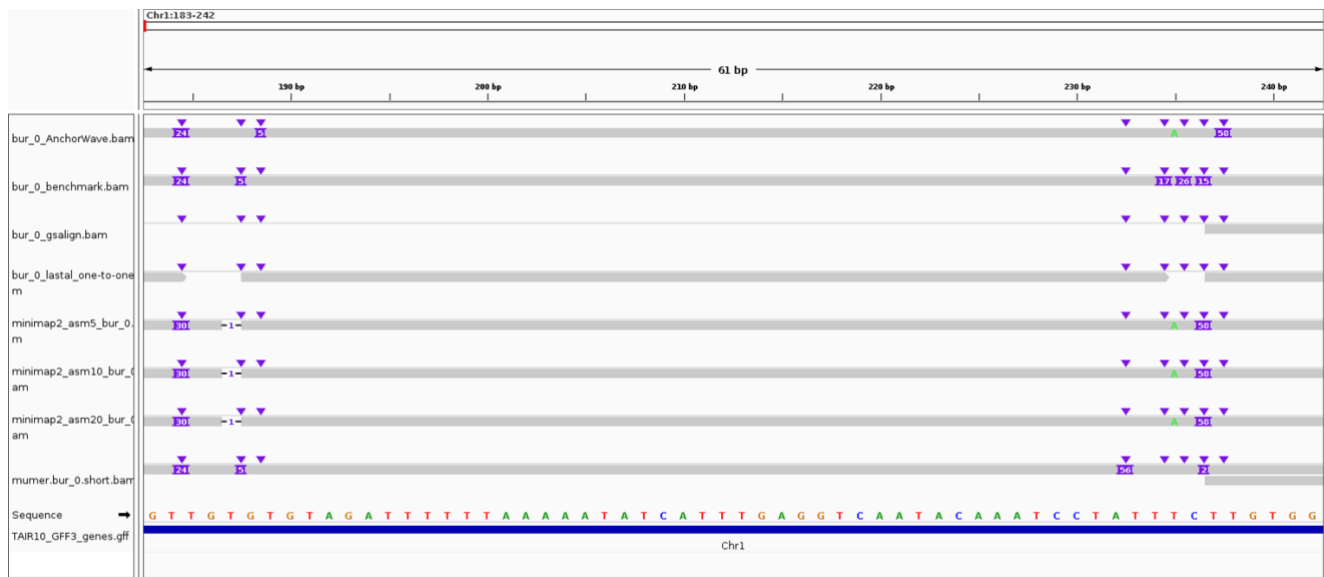


Fig. E2.

1) In the leading region of this window:

The LAST one-to-one approach and the GSalgn program did not generate alignment. All the other programs generated two while different INDELs. In the perspective of sequence alignment scoring, those alignments are equally optimized.

2) In the tailing region of this window:

The LAST one-to-one approach and the GSalgn program did not generate alignment. There are three INDELs in the benchmark alignment. AnchorWave and the three settings of minimap2 identified an SNP and an INDEL. Mummer identified two INDELs. In the perspective of sequence alignment scoring and the principle of parsimony, the alignment of AnchorWave and minimap2 are more optimized.

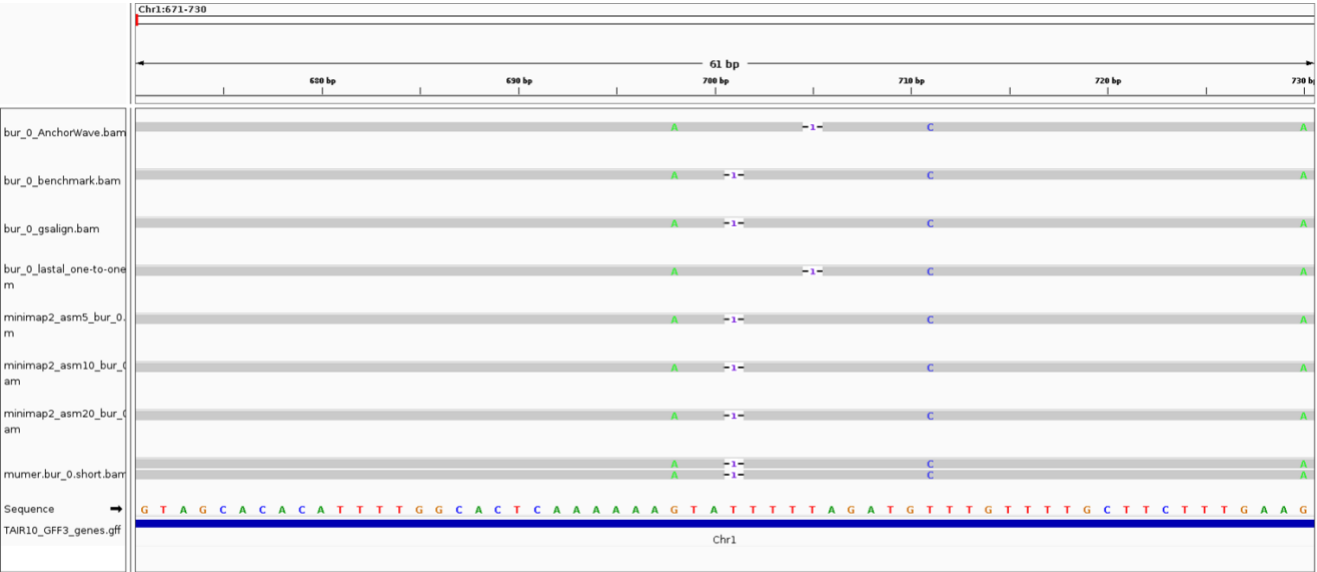


Fig. E3. All the programs generated a one base-pair deletion, although at different positions. In the perspective of sequence alignment scoring, all those alignments are equally optimized.

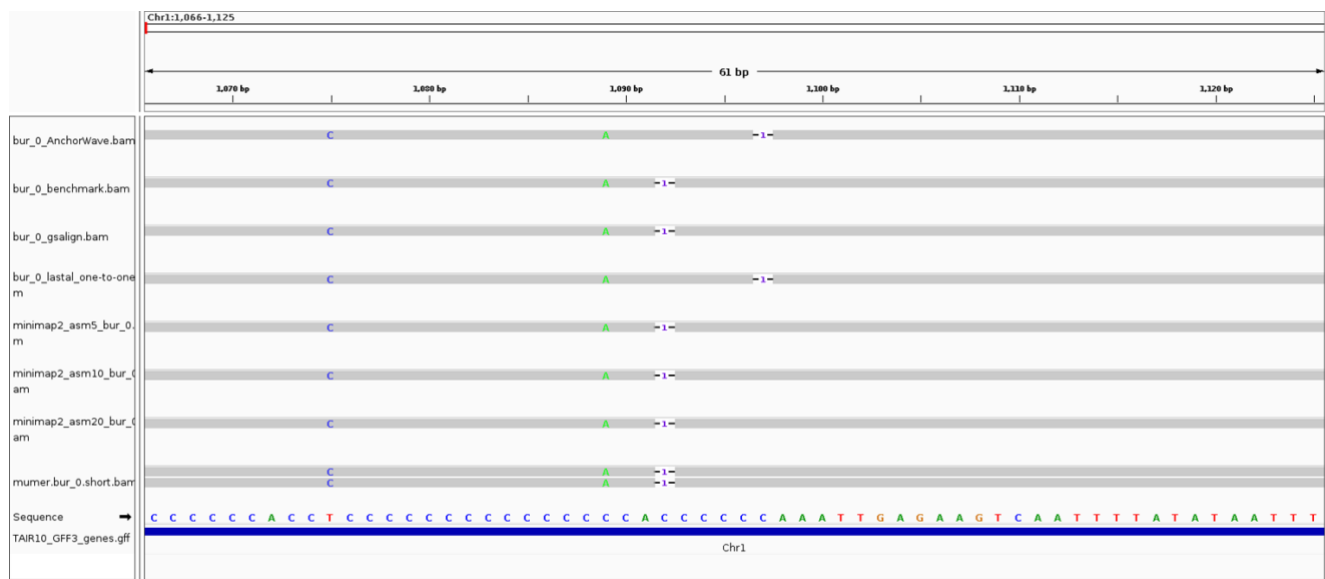


Fig. E4. All the programs generated a one base-pair deletion, although at different positions. In the perspective of sequence alignment scoring, all those alignments are equally optimized.

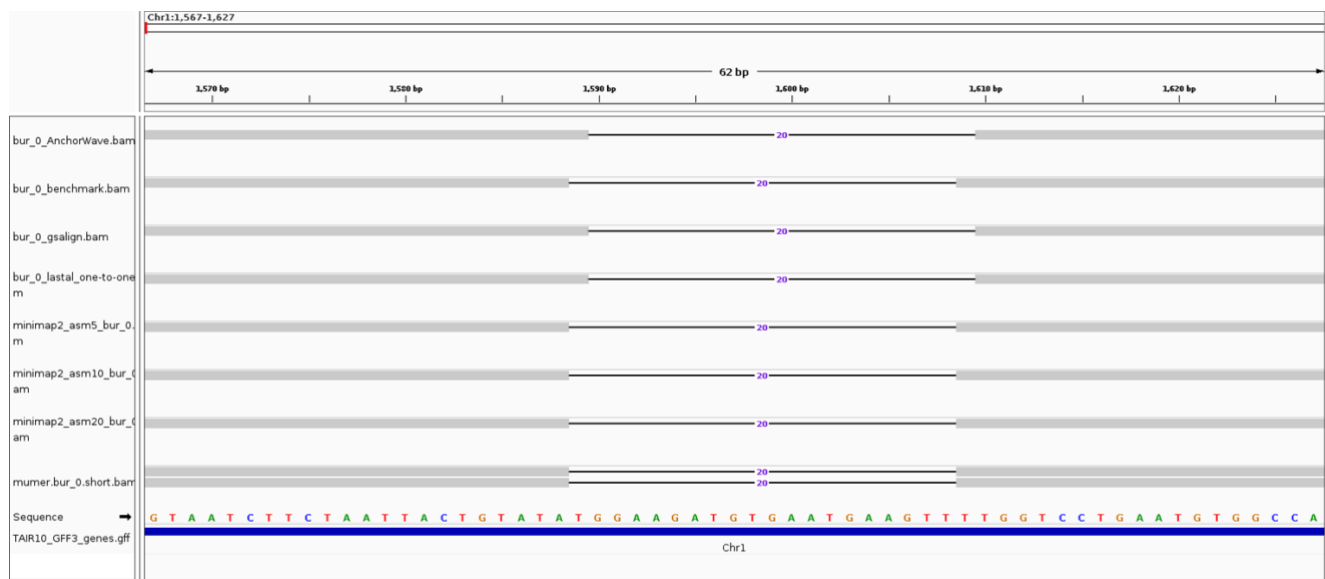


Fig. E5. All the programs generated a 20 base-pair deletion, although at different positions. In the perspective of sequence alignment scoring, all those alignments are equally optimized.

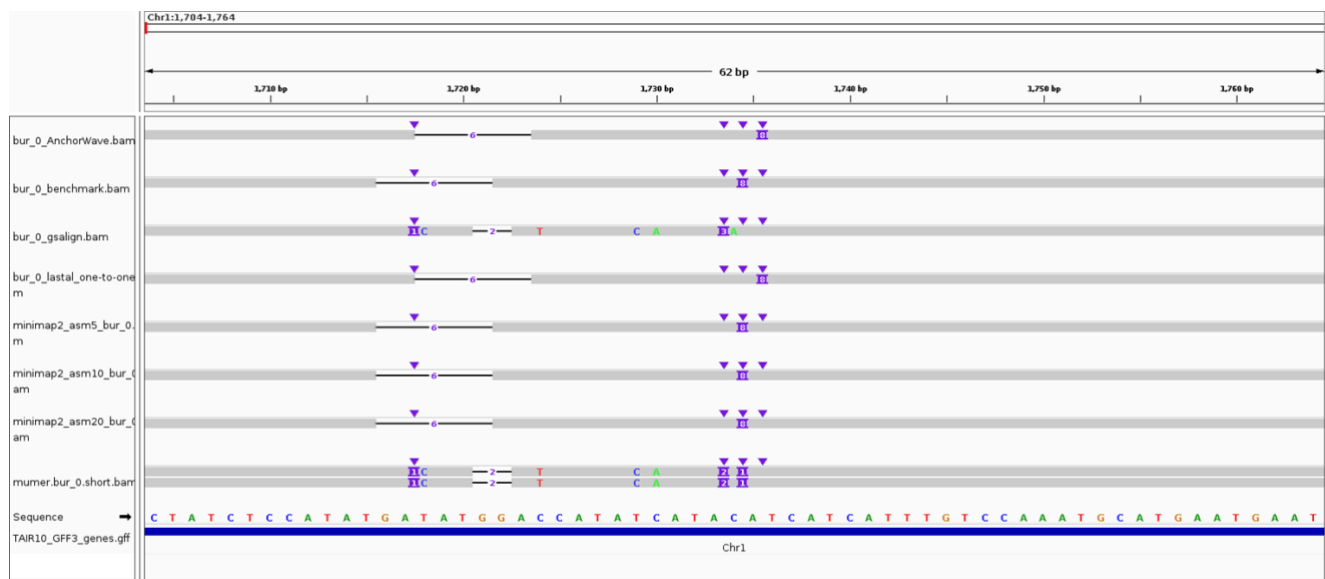


Fig. E6. AnchorWave, benchmark, LAST one-to-one, and the three settings of minimap2 generated a six base-pair deletion and an eight base-pair insertion, although at different positions. MUMmer4 and GSAIgen reported a larger number of variants. In the perspective of sequence alignment scoring and the principle of parsimony, MUMmer4 and GSAIgen generated less optimized sequence alignment.

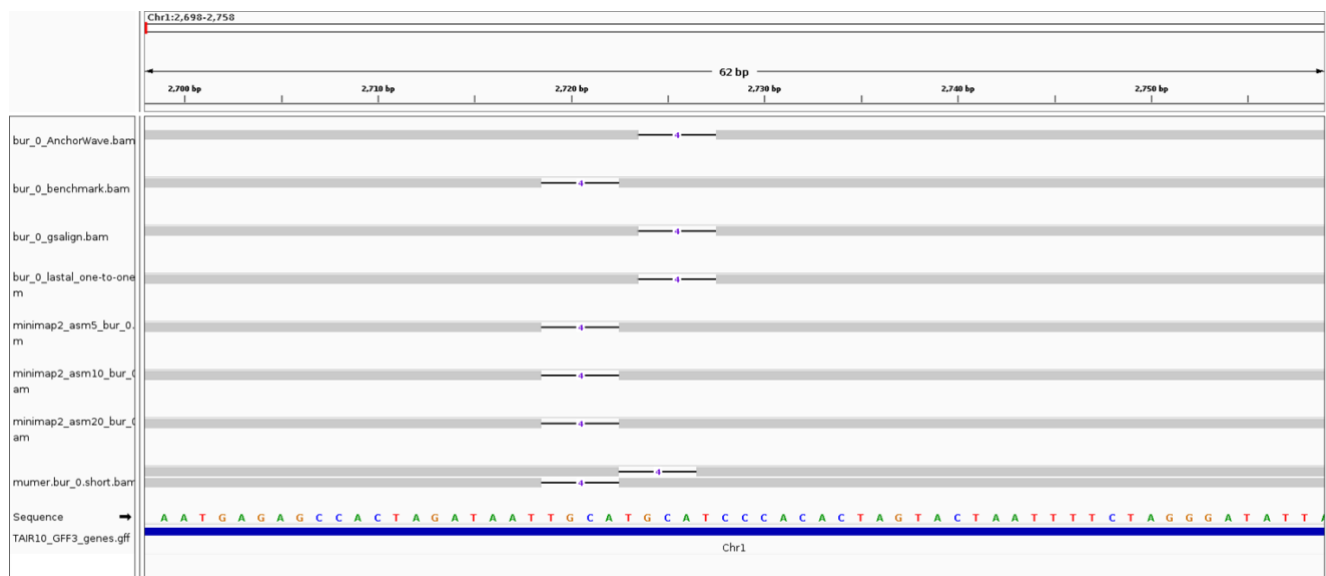


Fig. E7. All the programs generated a four base-pair deletion, although at different positions. In the perspective of sequence alignment scoring, all those alignments are equally optimized.

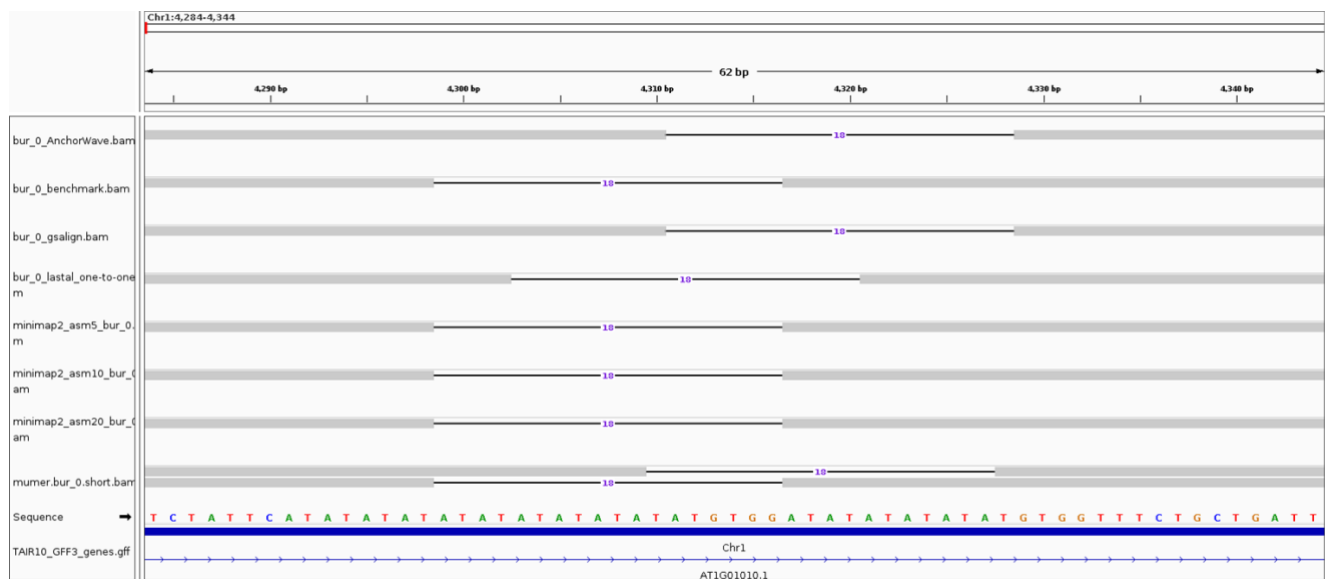


Fig. E8. All the programs generated an 18 base-pair deletion, although at different positions. In the perspective of sequence alignment scoring, all those alignments are equally optimized.

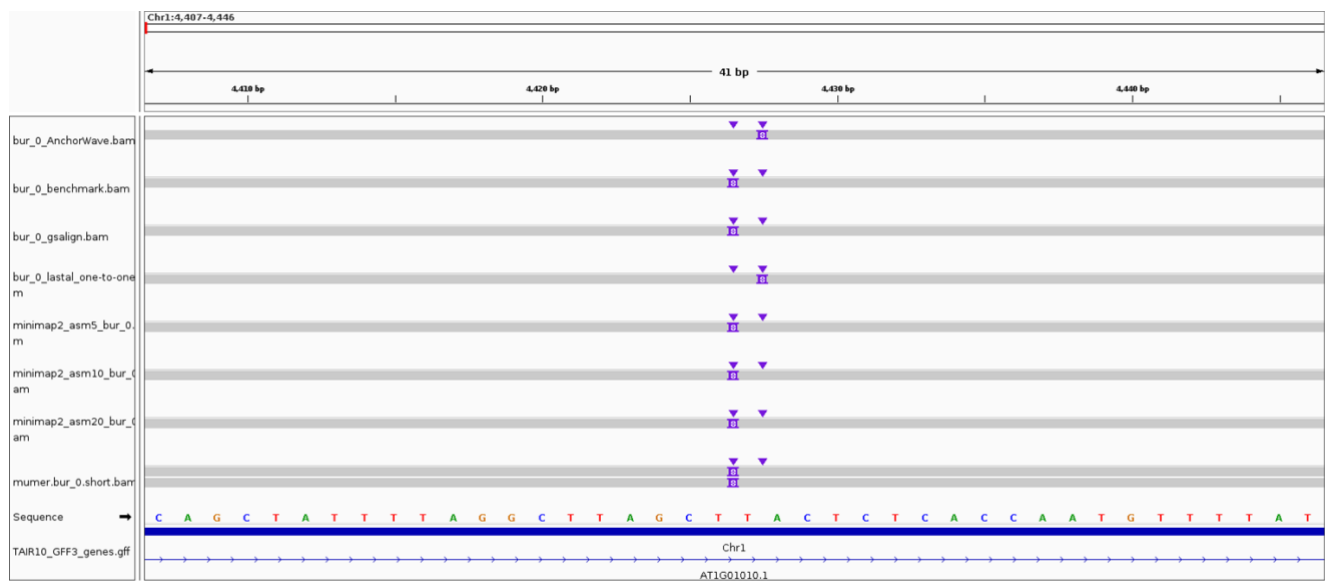


Fig. E9. All the programs generated an eight base-pair insertion, although at different positions. In the perspective of sequence alignment scoring, all those alignments are equally optimized.

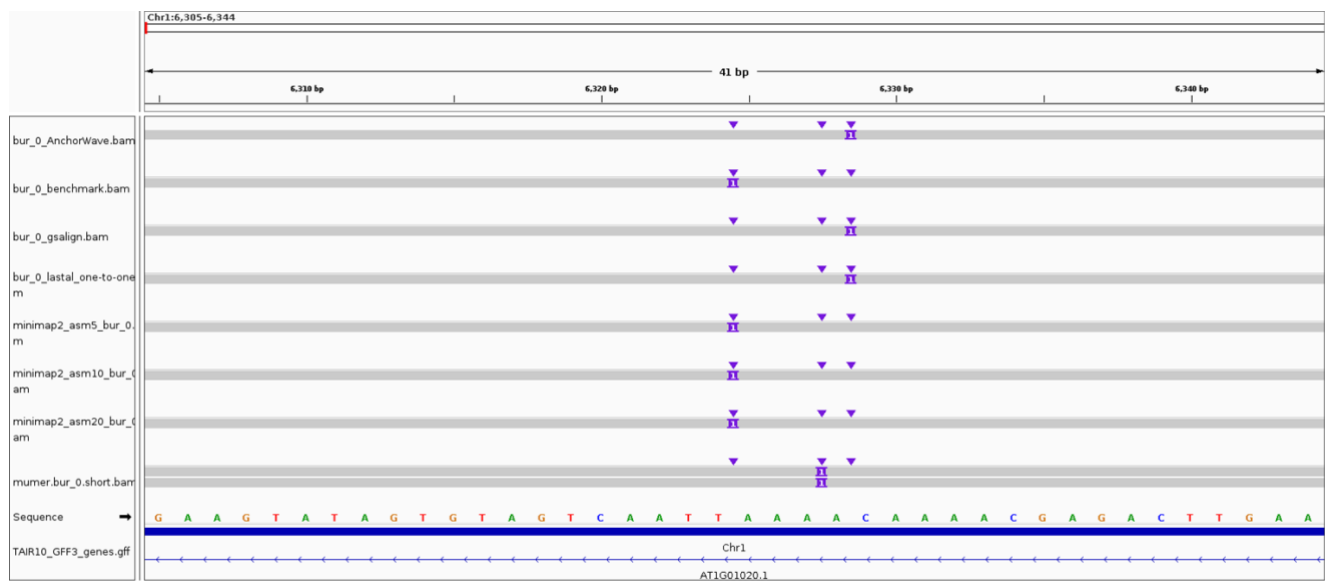


Fig. E10. All the alignment programs generated a one base-pair insertion, although at different positions. In the perspective of sequence alignment scoring, all those alignments are equally optimized.

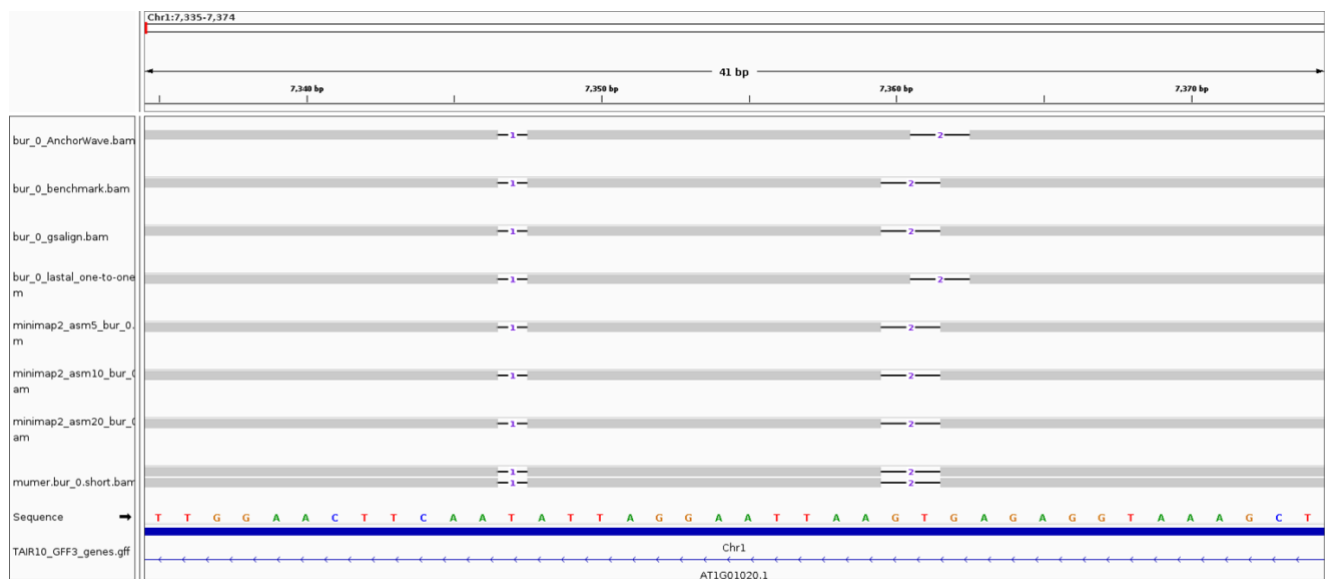


Fig. E11. All the programs generated a 1 base-pair deletion and a two base-pair deletion, although at different positions. In the perspective of sequence alignment scoring, all those alignments are equally optimized.

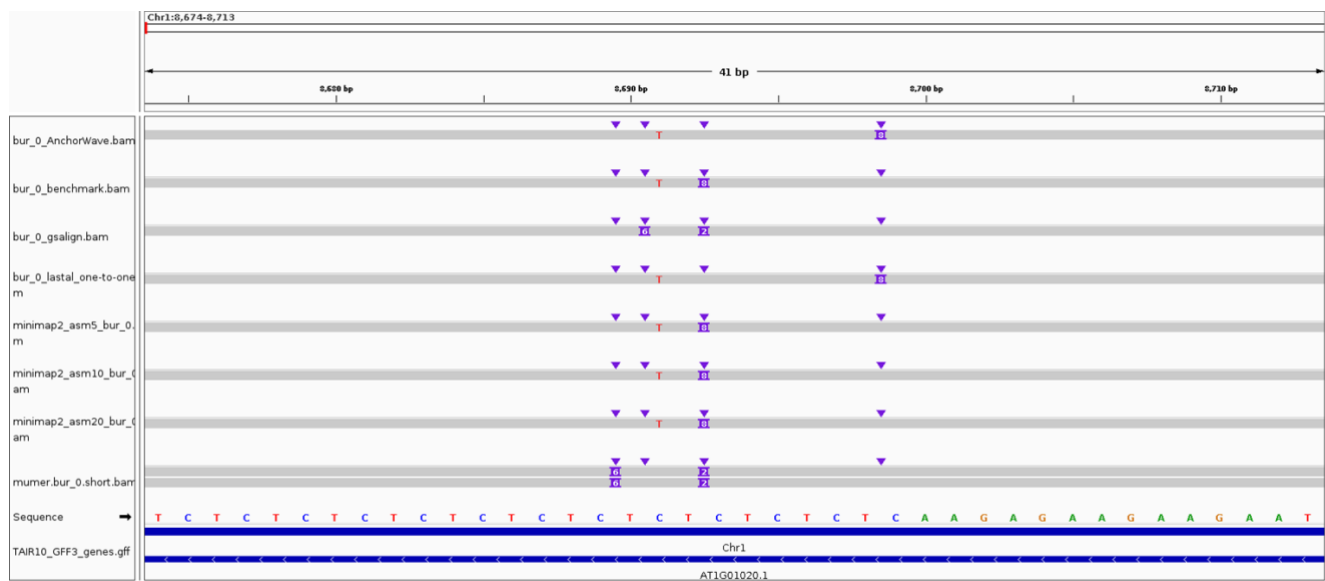


Fig. E13. Similar with the benchmark, AnchorWave, minimap2, LAST-one-to-one generated an SNP and an eight base-pair insertion. While GSAalign and MUMmer4 generated a 6 base-pair insertion and a 2 base-pair insertion.

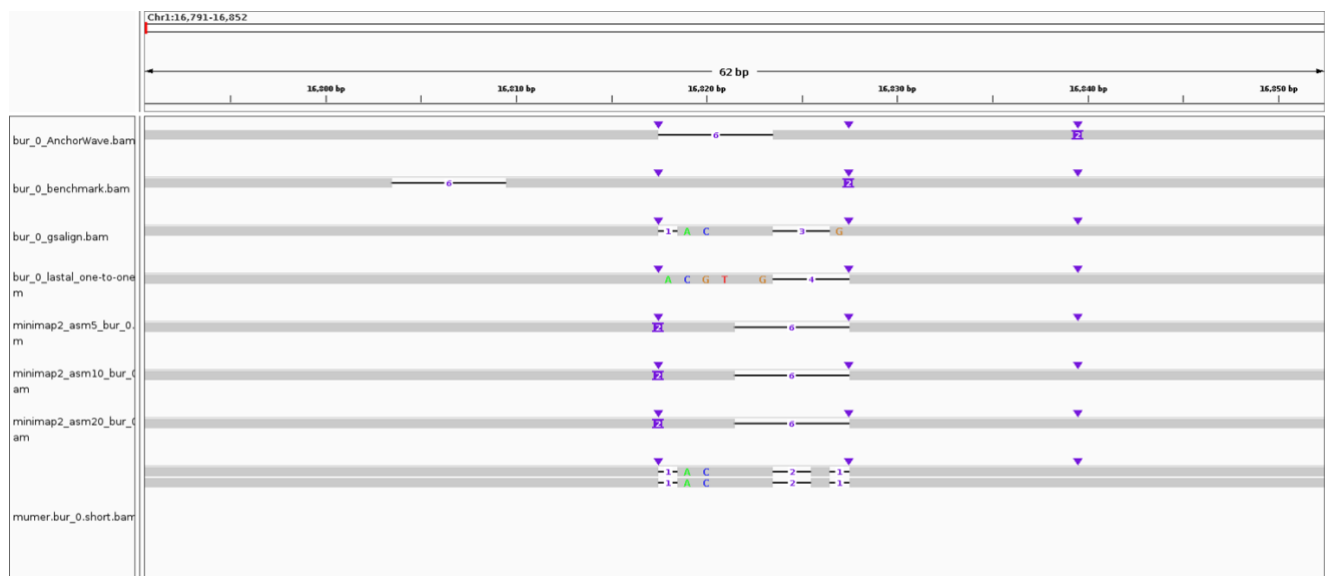


Fig. E14. Similar with the benchmark, AnchorWave and minimap2 generated a six base-pair deletion and a two base-pair insertion. While GSAIalign, MUMmer4 and LAST one-and-one generated larger number of variants.

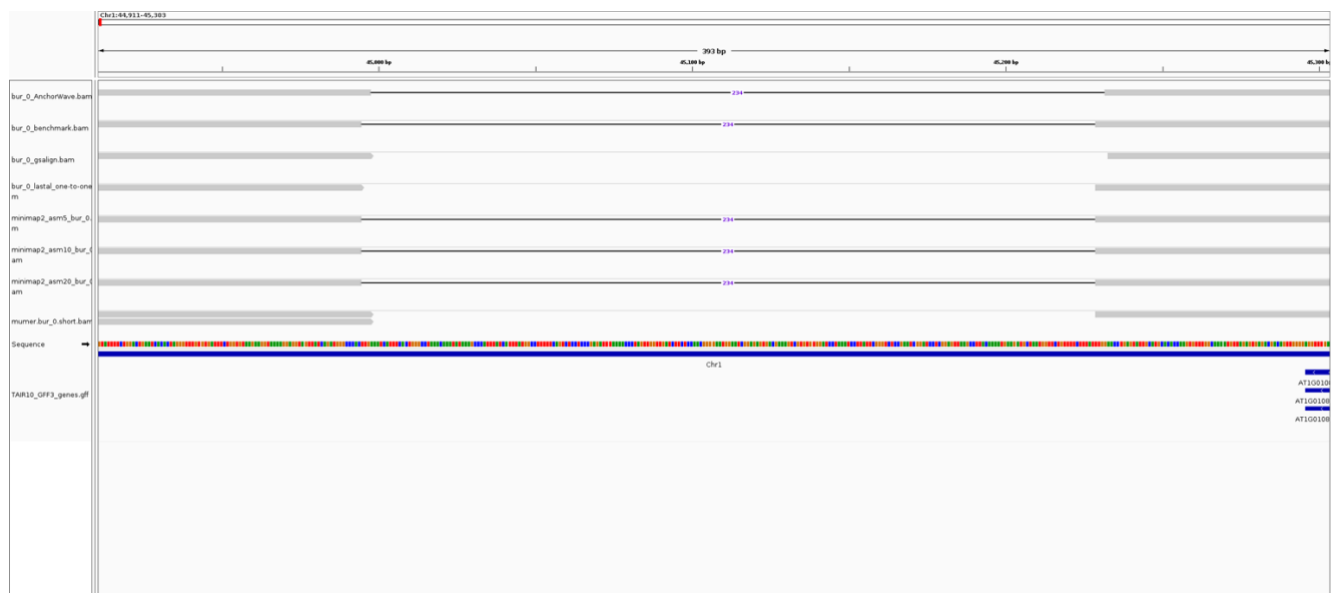


Fig. E16. Similar to the benchmark, AnchorWave and minimap2 generated a 234 base-pair deletion. While GSAalign, LAST one-to-one, and MUMmer4 failed to generate alignment across over in this region.



Fig. E17. Similar to the benchmark, AnchorWave, GSAIalign, MUMmer4, and minimap2 generated a 24 base-pair insertion and a one base-pair deletion. While LAST one-to-one generated three INDELs and many SNPs, in the perspective of sequence alignment scoring and the principle of parsimony, LAST one-to-one generated less optimized sequence alignment.

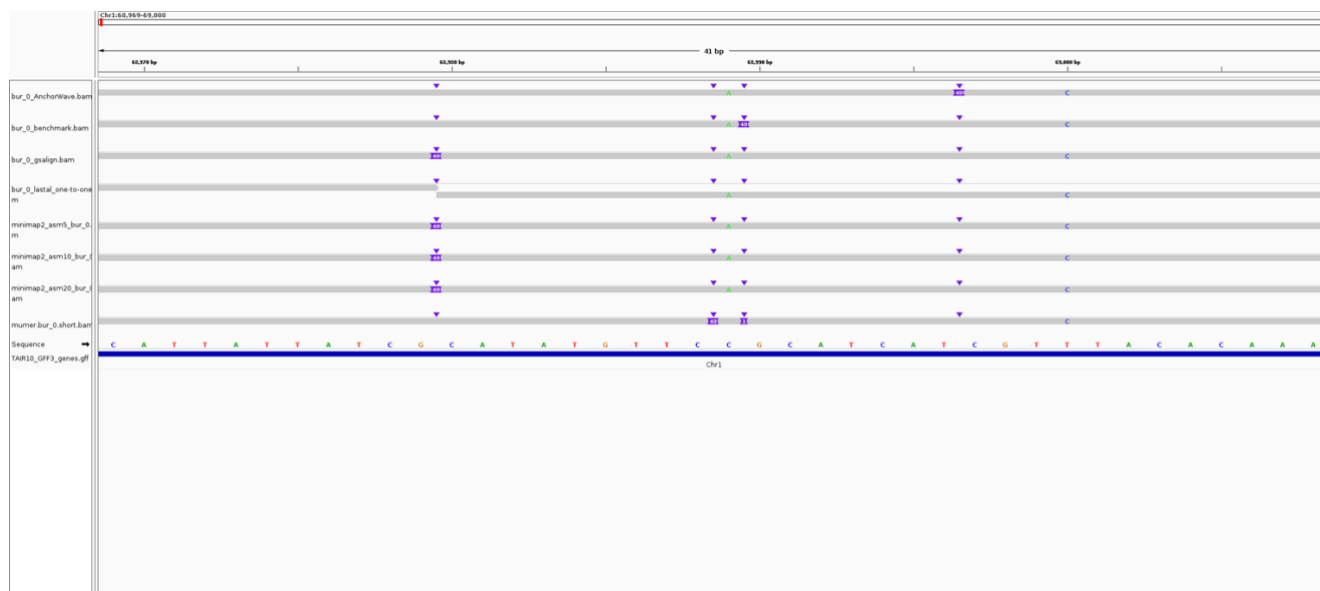


Fig. E18. Similar to the benchmark, AnchorWave, GSAIalign, and minimap2 generated a 49 base-pair insertion and two SNPs. While LAST one-to-one failed to align across this region. MUMmer4 generated a 48 base-pair insertion, a one base pair insertion, and an SNP.



Fig. E19. Different alignments were generated. In the perspective of sequence alignment scoring and the principle of parsimony, MUMmer4 generated the least optimized alignment.

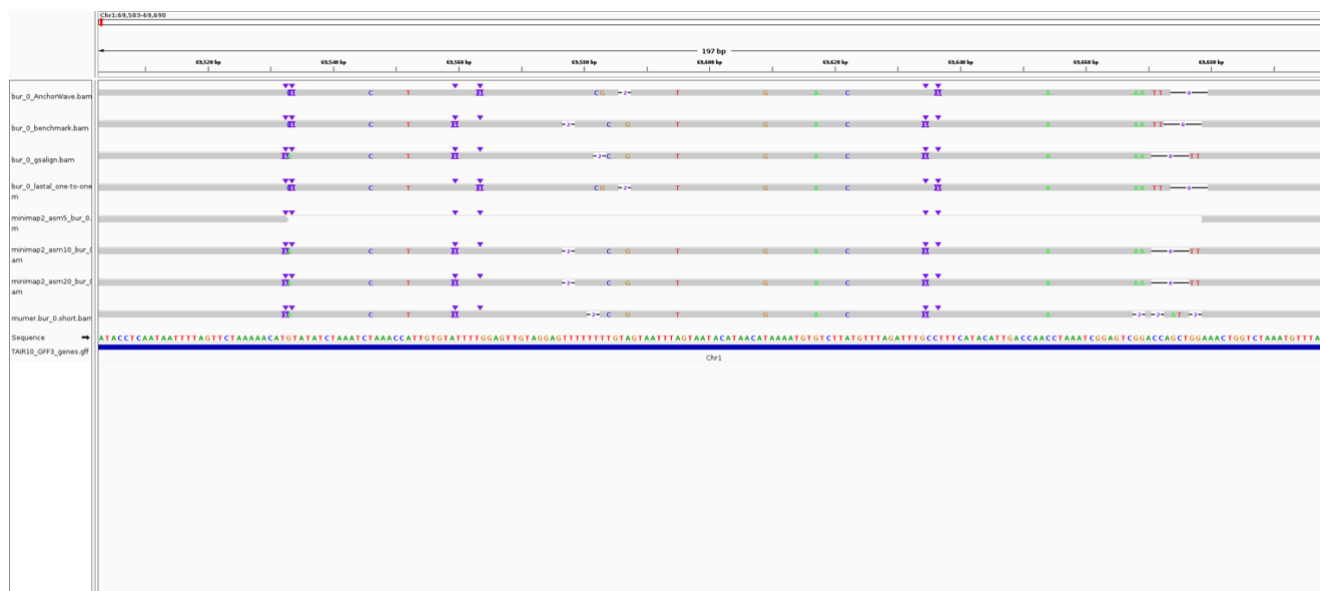


Fig. E20. MUMmer4 generated an alignment with more indels while fewer SNPs. Minimap2 with the asm5 setting did not generate alignment cross over this region. All the other programs generated the same number and type of variants, although at different positions.

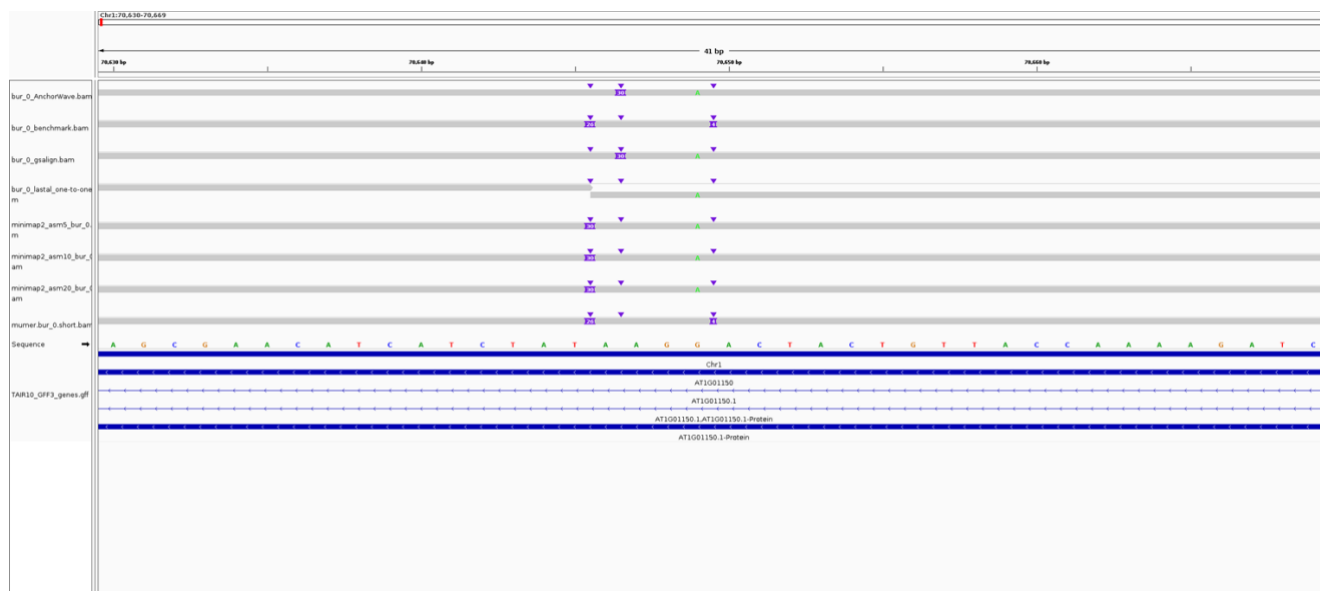


Fig. E21. LAST one-to-one did not generate alignment cross over this region. MUMmer4 generated alignment identical with the benchmark. AnchorWave, GSAIalign and minimap2 generated alignments with the same number and type of variants.

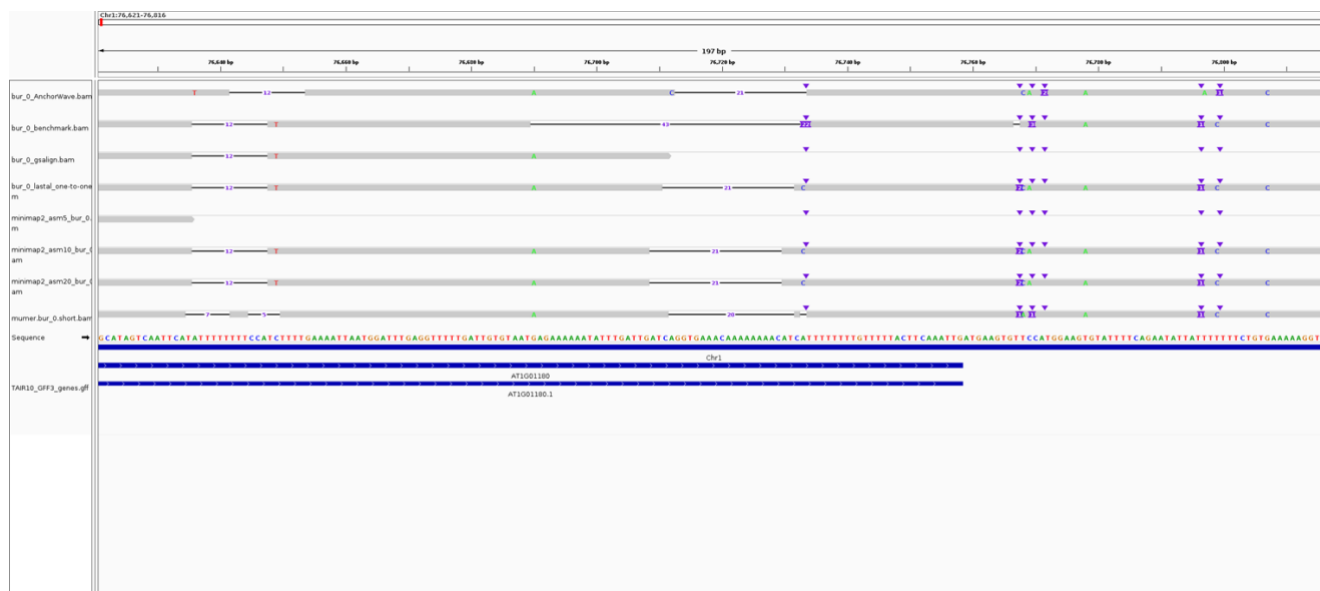


Fig. E22. In the middle of the window, the benchmark did not perform alignment. GSAalign and miniamp2 asm5 did not generate alignment cross over this region. AnchorWave, minimap2 asm10 and minimap2 asm20 generated alignments with the same number and type of variants. Comparing with AnchorWave, MUMmer4 generated the same number of variants, while more INDELs and fewer SNPs.

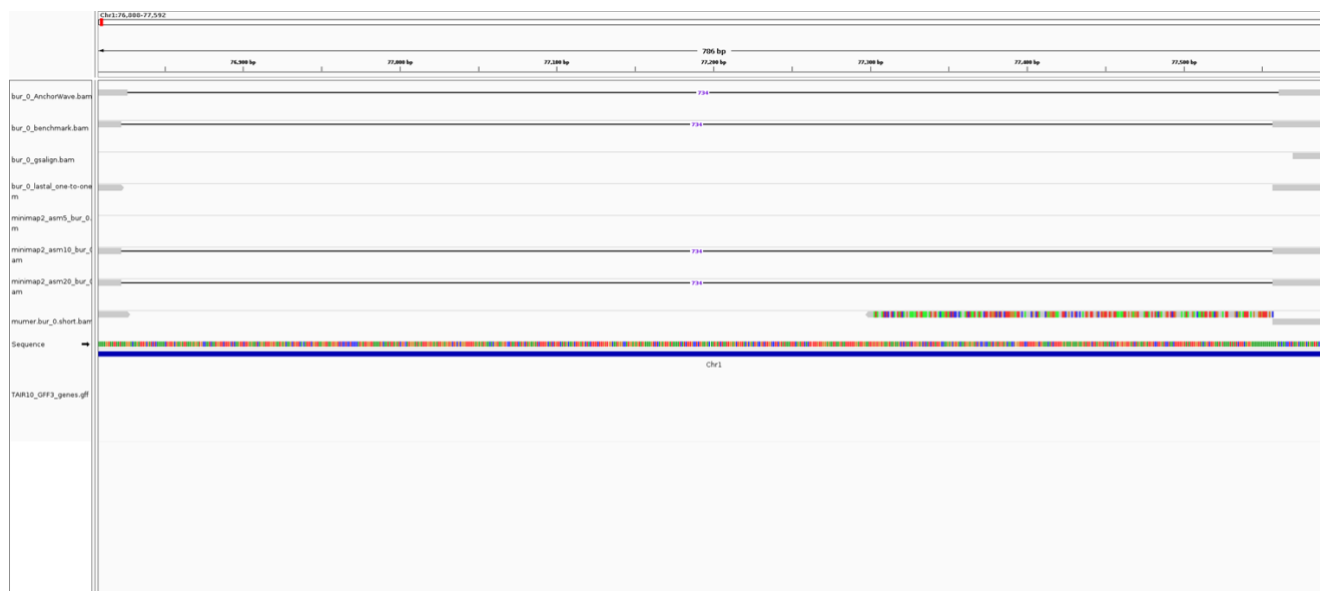


Fig. E23. AnchorWave, benchmark, minimap2 asm10, and minimap2 asm20 generated a 734 base-pair deletion. GSAalign, LAST one-to-one, minimap2 asm5, and MUMmer4 did not generate alignment cross over this region. MUMmer4 aligned a fraction of Chr5 to this region, generated unwanted alignment and many variants absent from the benchmark alignment.

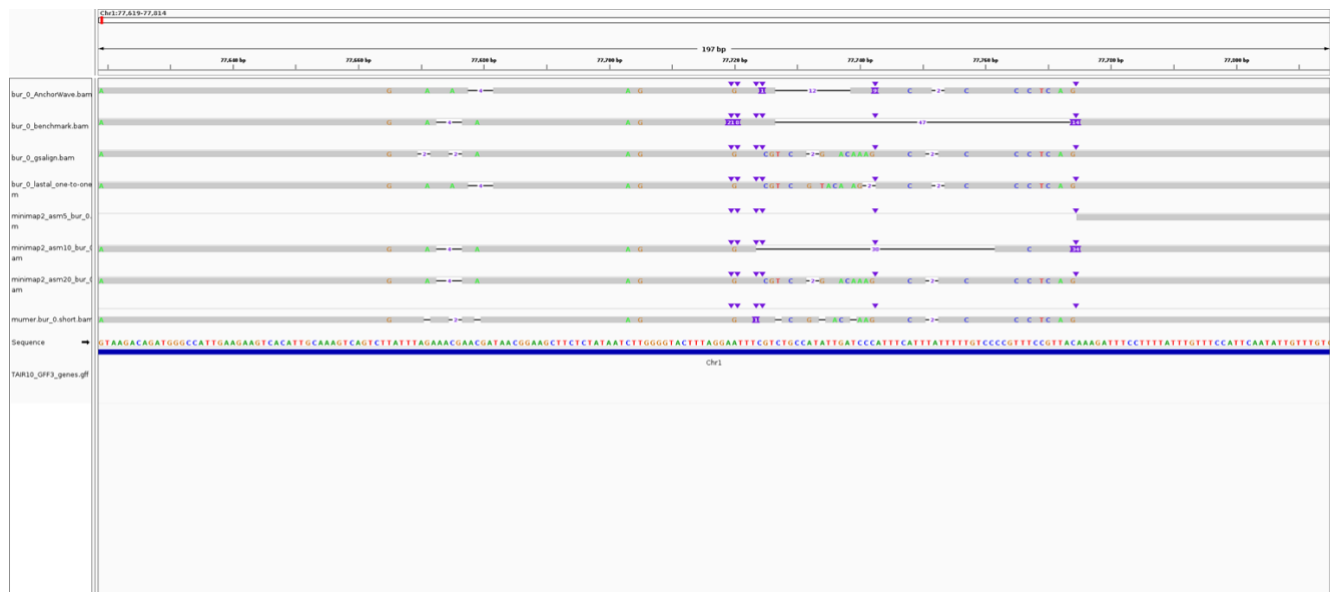


Fig. E24. The benchmark did not perform alignment for part of this region. miniamp2 asm5 did not generate alignment cross over this region. The other programs and settings generated different alignments.



Fig. E25. LAST one-to-one generated an alignment with different number of variants comparing with other programs.

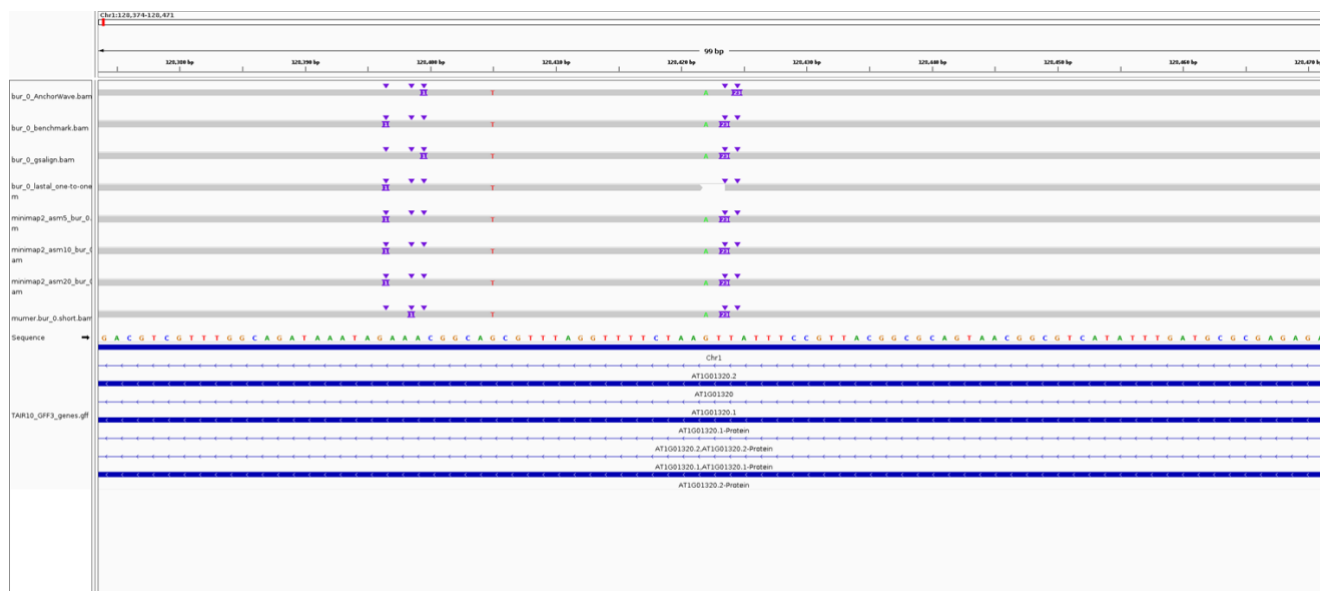


Fig. E26. LAST one-to-on failed to generate an alignment cross over this region. All the other programs and settings generated alignments with the same number and type of variants.

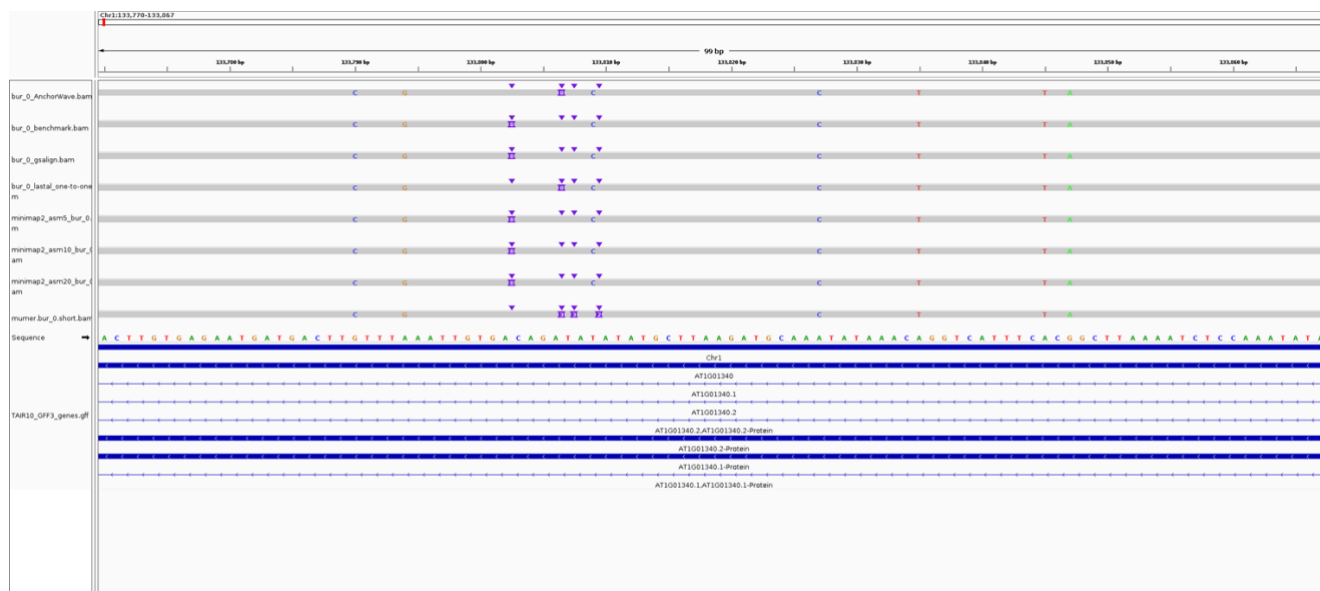


Fig. E27. MUMmer4 generated alignment with larger number of INDELs and also total number of variants. In the perspective of sequence alignment scoring and the principle of parsimony, MUMmer4 generated less optimized sequence alignment.



Fig. E29. Programs and settings generated different alignments.

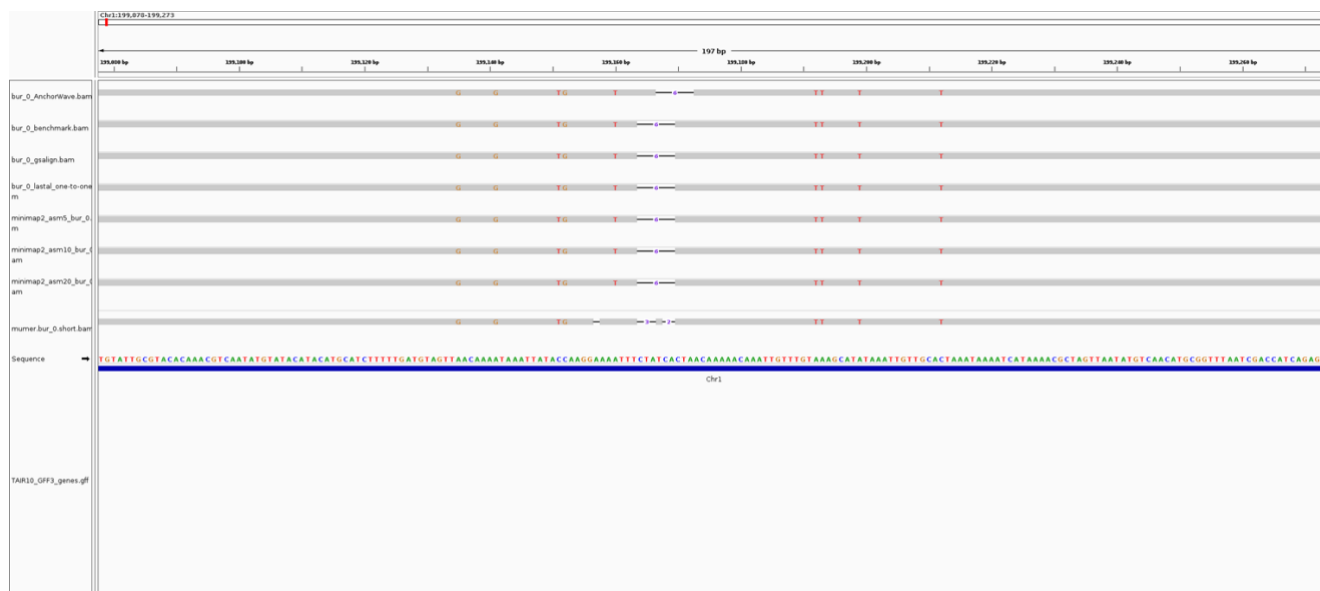


Fig. E30. With the principle of parsimony, MUMer4 generated less optimized alignment, due to more indels and total number of variants.