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/10.3389%2Ffgene.2019. https://dx.doi.org/10.3389%2Ffgene.2019. 01046, https://dx.doi.org/10.1371/journal.pgen.1007699) were skipped, due to they are so common.

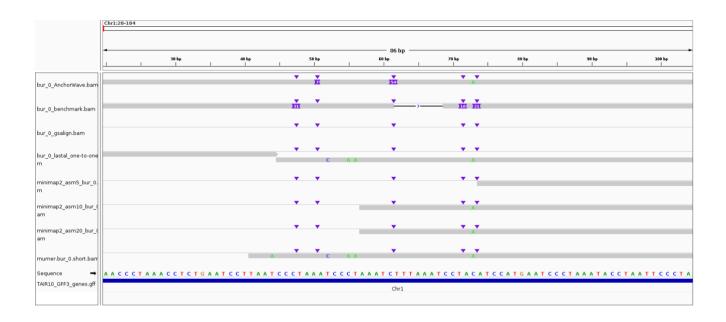


Fig. E1. In this window, AnchorWave identified two INDELs and an SNP. The benchmark reported 4 INDELs. While other programs did not cover the reference genome or query genome completely, likely due to the usage of local alignment strategy. In the perspective of sequence alignment scoring and the principle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

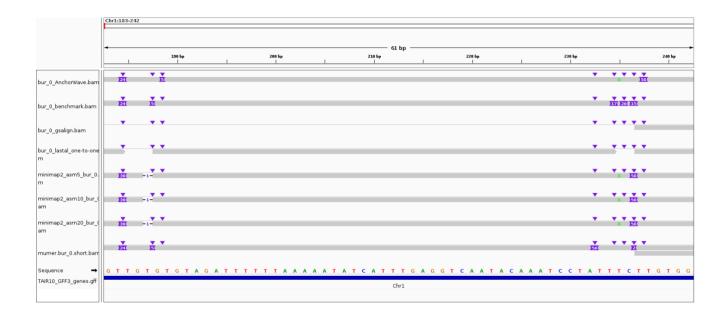


Fig. E2.

1) In the leading region of this window:

The LAST one-to-one approach and the GSAlign 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 GSAlign 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.

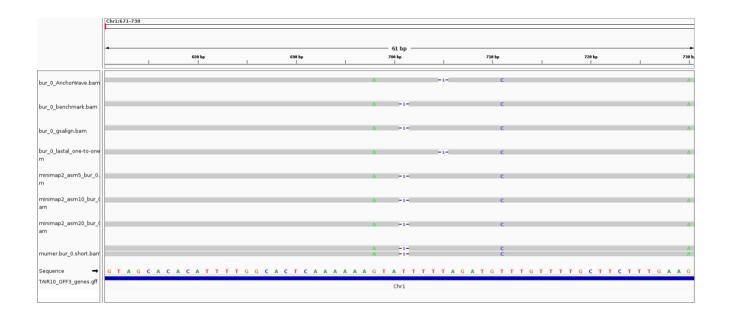


Fig. E3. All the programs generated a one base-pair deletion, although at different positions. In the perspective of sequence alignment scoring, all those alignments are equally optimized.

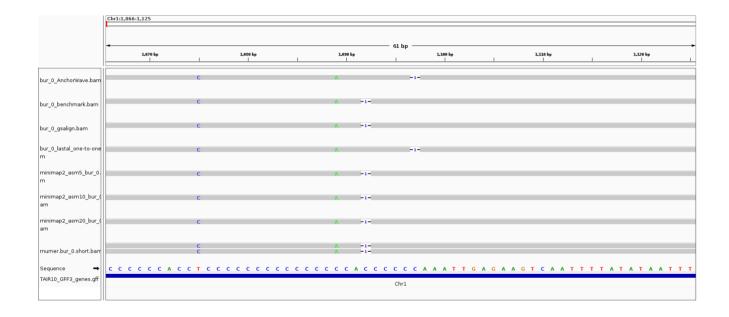


Fig. E4. All the programs generated a one base-pair deletion, although at different positions. In the perspective of sequence alignment scoring, all those alignments are equally optimized.

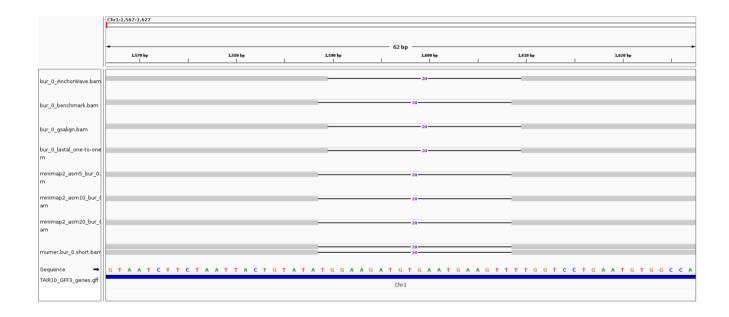


Fig. E5. All the programs generated a 20 base-pair deletion, although at different positions. In the perspective of sequence alignment scoring, all those alignments are equally optimized.

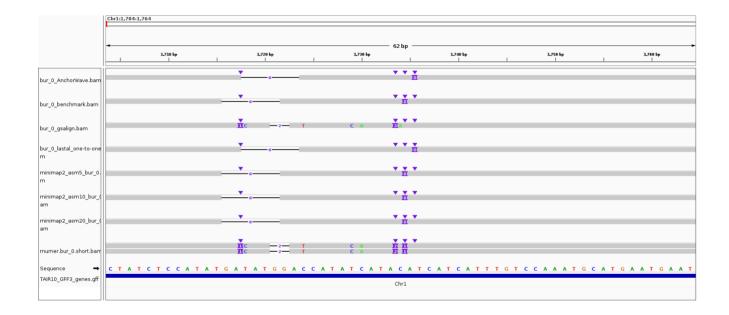


Fig. E6. AnchorWave, benchmark, LAST one-to-one, and the three settings of minimap2 generated a six base-pair deletion and an eight base-pair insertion, although at different positions. MUMmer4 and GSAlign reported a larger number of variants. In the perspective of sequence alignment scoring and the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

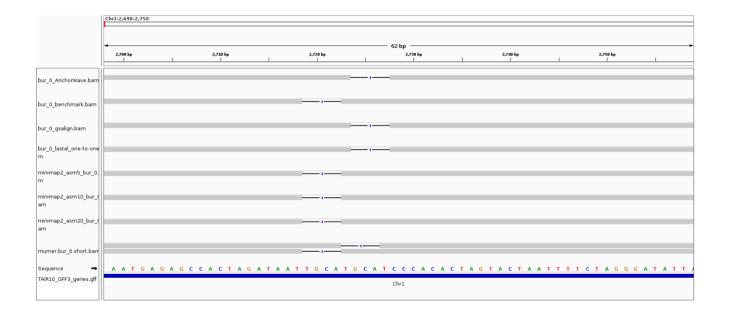


Fig. E7. All the programs generated a four base-pair deletion, although at different positions. In the perspective of sequence alignment scoring, all those alignments are equally optimized.

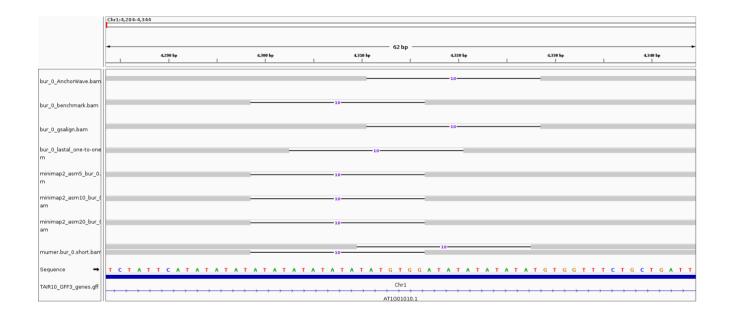


Fig. E8. All the programs generated an 18 base-pair deletion, although at different positions. In the perspective of sequence alignment scoring, all those alignments are equally optimized.

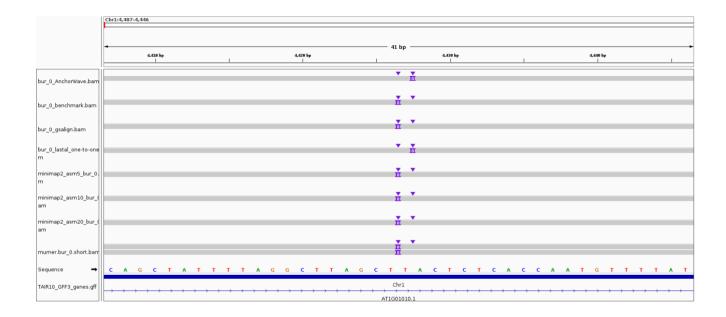


Fig. E9. All the programs generated an eight base-pair insertion, although at different positions. In the perspective of sequence alignment scoring, all those alignments are equally optimized.

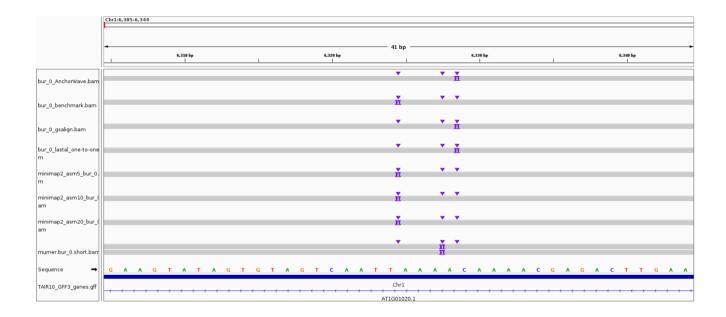


Fig. E10. All the alignment programs generated a one base-pair insertion, although at different positions. In the perspective of sequence alignment scoring, all those alignments are equally optimized.

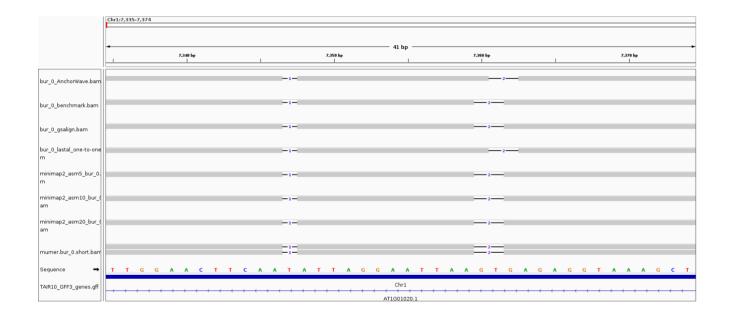


Fig. E11. All the programs generated a 1 base-pair deletion and a two base-pair deletion, although at different positions. In the perspective of sequence alignment scoring, all those alignments are equally optimized.

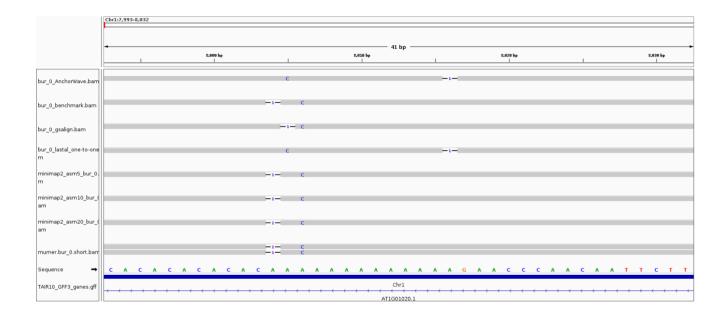


Fig. E12. All the programs generated a one base-pair deletion and an SNP, although at different positions. In the perspective of sequence alignment scoring, all those alignments are equally optimized.

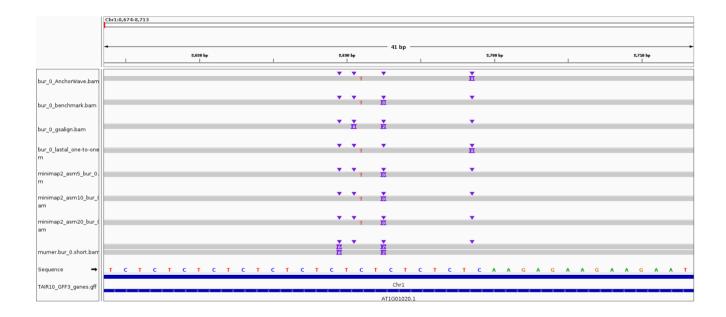


Fig. E13. Similar with the benchmark, Anchor Wave, minimap 2, LAST-one-to-one generated an SNP and an eight base-pair insertion. While GSAlign and MUMmer 4 generated a 6 base-pair insertion and a 2 base-pair insertion.

	Chr1:16,791-16,852							
	1	16,200 bp	ı	16,010 bp	62 bp —— 16,820 bp	16,830 bp	16,040 bp	16,350 bp
bur_0_AnchorWave.bam					6——	▼	¥ El	
bur_0_benchmark.bam			6	_	•	¥ El	•	
bur_0_gsalign.bam					-1- A C -	—3— G	*	
bur_0_lastal_one-to-one m					A C G T G		•	
minimap2_asm5_bur_0. m					ř –	-6	*	
minimap2_asm10_bur_0					<u>*</u>	-6	•	
minimap2_asm20_bur_0 am					<u>*</u>	-6	•	
					-1- A C	2= 11	•	
mumer.bur_0.short.bam								

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

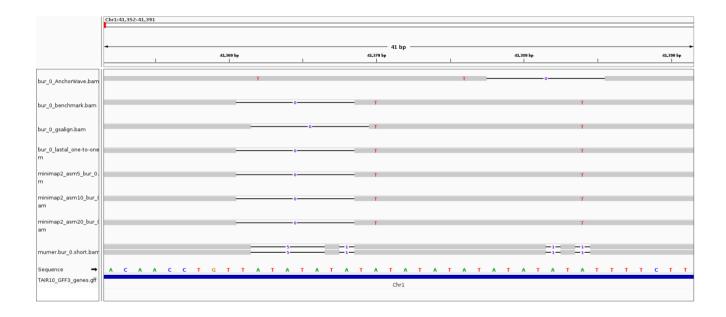


Fig. E15. Similar to the benchmark, Anchor Wave, GSAlign, LAST one-to-one, and minimap2 generated an eight base-pair deletion and two SNPs. While MUMmer4 generated four deletions. In the perspective of sequence alignment scoring and the principle of parsimony, MUMmer4 generated less optimized sequence alignment.



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



Fig. E17. Similar to the benchmark, Anchor Wave, GSAlign, 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.

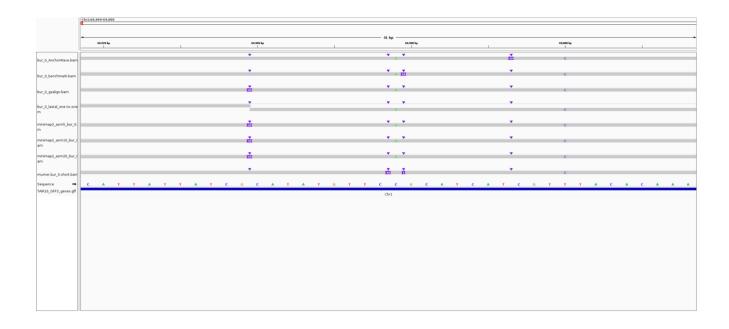


Fig. E18. Similar to the benchmark, Anchor Wave, GSAlign, and minimap 2 generated a 49 base-pair insertion and two SNPs. While LAST one-to-one failed to align across this region. MUMmer 4 generated a 48 base-pair insertion, a one base pair insertion, and an SNP.

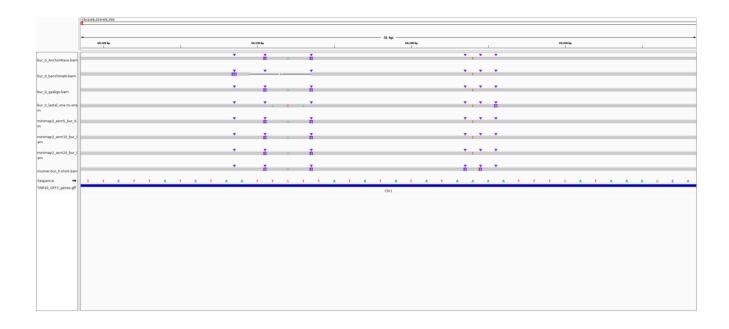


Fig. E19. Different alignments were generated. In the perspective of sequence alignment scoring and the principle of parsimony, MUMmer4 generated the least optimized alignment.

	Chr1:69,503-69,690								
	63,520 bp	63,540 kp	69,560 bp	63,500 hp	197 bp 63,600 bp	63,620 bp	63,640 kg	63,660 kp	6 4 903,63
bur_0_AnchorWave.bam		T c	, , ,	CG -2-	T G	A C	¥ ¥	A	AA TT :—
bur_0_benchmark.bam		T .	T T	-2- C G	T G	Α C	Ĭ,	•	AA TT—
bur_0_gsalign.bam		m c	T N	-2-C G	T G	A C	Ĭ,	^	AA
bur_0_lastal_one-to-one m		T	, , ,	CG -2-	T G	Α C	* <u>*</u>	^	AA TT
minimap2_asm5_bur_0. m		**	• •				* *		
minimap2_asm10_bur_0 am		n c	T N T	F2= C G	T G	Α	ň *	^	AA
minimap2_asm20_bur_6 am		n c	, , , ,	F2F C G	T G	A C	ň *	A	AA :
mumer.bur_0.short.ban		n c	, ř. ř.	-2- C G	T G	А С	, v	٨	-22- AT -2-
Sequence TAIR10_GFF3_genes.gff	ATACCTCAATAATTTTAGTTCTAAAAACA	TGTATATCTAAATCTAAAC	CATTGTGTATTTTGGAGTTGT	AGGAGTTTTTTTTGTAGTAATTT	AGTAATACATAACATAAAA	ATGTGTCTTATGTTTAGA	TTTGCCTTTCATACATTGAC	CAACCTAAATCGGAG	TCGGACCAGCTGGAAACTGGTCTAAATGTTTAA

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.

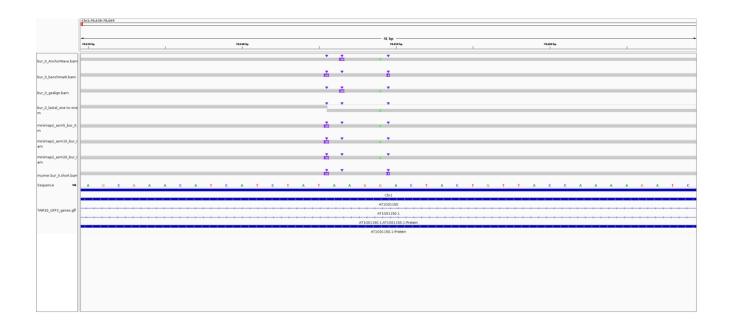


Fig. E21. LAST one-to-one did not generate alignment cross over this region. MUMmer4 generated alignment identical with the benchmark. AnchorWave, GSAlign and minimap2 generated alignments with the same number and type of variants.

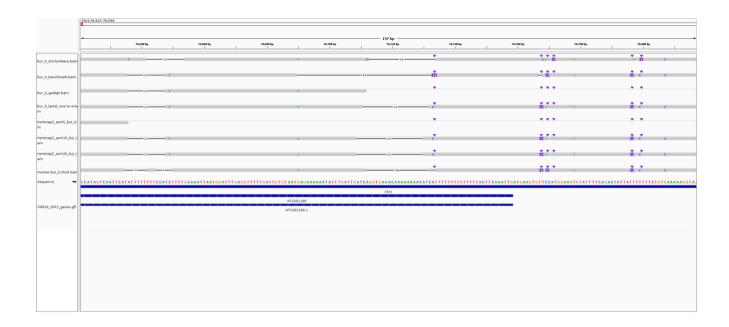


Fig. E22. In the middle of the window, the benchmark did not perform alignment. GSAlign and miniamp2 asm5 did not generate alignment cross over this region. AnchorWave, minimap2 asm10 and minimap2 asm20 generated alignments with the same number and type of variants. Comparing with AnchorWave, MUMmer4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

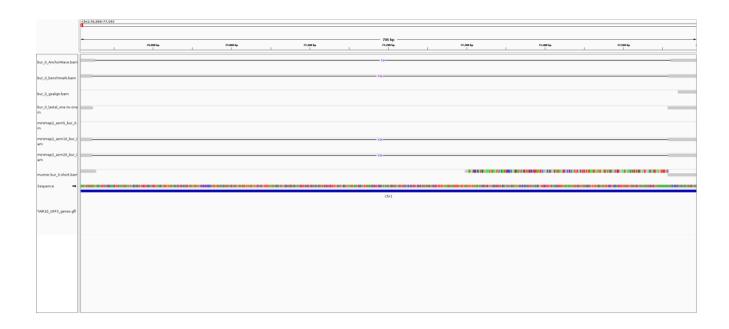


Fig. E23. AnchorWave, benchmark, minimap2 asm10, and minimap2 asm20 generated a 734 base-pair deletion. GSAlign, LAST one-to-one, minimap2 asm5, and MUMmer4 did not generate alignment cross over this region. MUMmer4 aligned a fraction of Chr5 to this region, generated unwanted alignment and many variants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

	Qualitic FF4 Life AFT
	99 bp
bur_0_AnchorWave.bam	, v v
bur_0_benchmark.bam	T T T A E
bur_0_gsalign.bam	* * <u>*</u> , <u>*</u> *
bur_0_lastal_one-to-one	<u> </u>
minimap2_asm5_bur_0.	* ** · · · · · · · · · · · · · · · · ·
minimap2_asm10_bur_0	i vv
minimap2_asm20_bur_0	T T T A M T
mumer.bur_0.short.ban	v řív
Sequence -	GA COT COTT TO G CAGATAAATAGAAACG G CAG COTTTAG GTTTT CTAAGTTATT CCGTTACG CG CAGTAACG G COTCATATTTGATG CG CO COAGAGA
	Ori A10013022
	A10(39)26
TAIR10_GFF3_genes.gff	AT061393.1
	AT1001320 J-Peten AT1001302 AT1001303 - Jetsein
	AT 001326 LAT 001326 1 - Insteam
	ATIO01120-2-Protein

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.

	Chr1:133,770-133,86	7																
	•																	
	•							99 bp —										
		131,71	00 bp	233	798 lp		333,000 bp		133,828 bp	133,620 bp	130	L230 Sp	233,840 %	•	333,050 kp		333,060 bp	
						6		¥ ¥			,			· ·				
bur_0_AnchorWave.bam																		
bur_0_benchmark.bam					С	G	¥		C C		С	т		т	A			
00.00.000																		
bur_0_gsalign.bam					С	G	ř	• •	c		C	T		T	A			
								* *										
bur_0_lastal_one-to-one m					С	G		n	C		c	Ť		T	A			
minimap2_asm5_bur_0.						6	×							,				
m																		
minimap2_asm10_bur_0					С	G		• •	c		c	т		T	A			
am																		
minimap2_asm20_bur_0 am					С	G	ň	* *	c		c	T		T	٨			
							•	ž ž										
mumer.bur_0.short.bam					С	G		nn	Ħ		c	Ť		T	A			
Sequence →	ACTTGT	GAGAA	TGATG	ACTT	G T T T	A A A T T	GTGAC	GATA	TAT	A T G C T T A A G A T G		T A A A C A	GGTCA	T T C A C	GGCTTA	A A A T C	T C C A A	A T A T A
										Chrl								
										AT1G01340								
										AT1G01340.1								
TAIR10_GFF3_genes.gff										AT1G01340.2								
										AT1G01340.2.AT1G01340.2-Protein	,							
										AT1G01340.2-Protein								
										AT1G01340.1-Protein								
										AT1G01340.1,AT1G01340.1-Protein	, , , , , ,							

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.

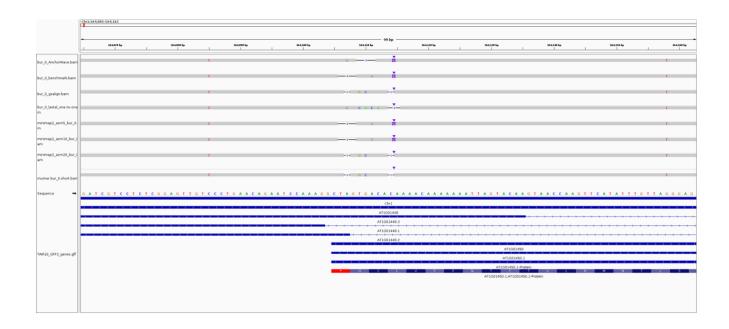


Fig. E28. Programs and settings generated different alignments.



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.