

Improved visualisation of GS20 traces in GAP4

Bernd Senf^{a*}, James Bonfield^b, Roman Siddiqui^a, Stefan Taudien^a and Matthias Platzer^a

- a) Genome Analysis, Leibniz Institute for Age Research, Fritz Lipmann Institute (FLI)
 Jena, Germany *) bsenf@fli-leibniz.de
- b) Sanger Centre, Wellcome Trust Genome Campus, Hinxton Cambridge, CB10 1SA, UK

Assembling GS20 sequences with runPhoenix results in ACE files, commonly visualised by different tools, e.g. Lasergene, Consed or GAP4 [1]. For GAP4, conversion of the ACE file by Phrap2caf and caf2gap [2] produces a GAP4 alignment (Fig. A) with three major drawbacks:

- Base confidence values are not available,
- Sequence parts that were excluded by the assembly program ("hidden data") can not be retrieved,
- The pyrograms ("traces") in their standard flowgram format (sff) can be visualized by a

mouseclick, but the trace position is shifted with respect to the appropriate position in the assembly.

To overcome these shortcomings, hampering the inspection and evaluation of GS20 sequence assemblies, we developed two programs **(sff_dump.c and roche454ace2caf.pl)**. Running these tools before starting caf2gap (Fig. B) results in GAP4 alignments:

- visualizing the entire sequence, including "hidden data", in the contig editor,
- assigning base confidence values to all read positions in the contig editor,
- visualizing the trace data at their correct positions

Α	B	runPhoenix => EIK123a.sff,EIK123b.sff	
runPhoenix => EIK123a.sff,EIK123b.sff		runAssembly, runMapping => 454Contigs.ace	
runAssembly, runMapping => 454Contigs.ace		sff_dump –f EIK123a.fna –q EIK123a.qual EIK123a.sff	
phrap2caf - newace => 454Contigs.caf		sff_dump –f EIK123b.fna –q EIK123b.qual EIK123b.sff	New C- and Perl based
		$racha454aac2aaf r1 \rightarrow 454Canting aaf$	programo

caf2gap –project 454 –version 0 < 455Contigs.caf

gap4.12 454.0

roche454ace2caf.pl => 454Contigs.caf

caf2gap –project 454 –version 0 < 455Contigs.caf





_						Co	ntig	Editor:	-12	274 EIK14	49E01	AIKY6				•
Cons	1 ;		Qual	1	ŧ		Insert	Edit Moc	les >>	📕 Cutoffs	Undo	Next Search	Commands >>	Settings >>	Quit	Help >>
<<	<		>	>	>	\triangleleft										
; +1; -1; -1; -4; +2; +4; +4; -10 -2; -2; +4; +10 Base	777 E 256 E +99 E -38 E 167 E 169 E 392 E 146 E 124 E 124 E 219 E 339 E 219 E 339 E 219 E 339 E 219 E 339 E 219 E 339 E 219 E 339 E 219 E	IK: IK: IK: IK: IK: IK: IK: IK: IK: IK:	149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(149E(1)381)2A6)381)2A6)2A6)2A6)2A6)2A6)2A6)2A6)2A6	A7X 0WK U4A F5E F5E EM0 5W5 EM0 5W5 EM0 5W5 EM0 5W5 KKS DHW DHW DHW DHW DHW DHW DHW DHW DHW S2T CLN XX- Prob	TG* TG* TG* TG* TGG TG* TG*	580 **AGAG **AGAG **AGAG **AGAG GGAGAG **AGAG **AGAG **AGAG **AGAG ity 1.01) 5 CAAAGAAGA CAAAGAAGA CAAAGAAGA CAAAGAAGA CAAAGAAGA CAAAGAAGA CAAAGAAGA CAAAGAAGA CAAAGAAGA CAAAGAAGA CAAAGAAGA CAAAGAAGAAGA CAAAGAAAGAAGA CAAAGAAAGAAGA	190 ITATCTI ITATCTI ITATCTI ITATCTI ITATCTI ITATCTI ITATCTI ITATCTI ITATCTI ITATCTI ITATCTI		610 GAGATTI GAGATTI GAGATTI GAGATTI GAGATTI GAGATTI GAGATTI GAGATTI) 620 *GATITCAT*CA *GATITCAT*CA *GATITCAT*CA *GATITCAT*CA *GATITCAT*CA *GATITCAT*CA *GATITCAT*CA *GATITCAT*CA *GATITCAT*CA *GATITCAT*CA *GATITCAT*CA CAT*CA T*CA	630 AAAAA*TTTG AAAAA*TTTGCATO AAAAA*TTTGCATO AAAAA*TTTGCATO AAAAA*TTTGCATO AAAAA*TTTGCATO AAAAA*TTTGCATO AAAAA*TTTGCATO AAAAA*TTTGCATO AAAAA*TTTGCATO AAAAA*TTTGCATO AAAAA*TTTGCATO	640 CCA CCATTTATGTTTA CCATTTATGTTTA CCATTTATGTTTA CCATTTATGTTTA CCATTTATGTTTA CCATTTATGTTTA CCATTTATGTTTA CCATTTATGTTTA CCATTTATGTTTA CCATTTATGTTTA CCATTTATGTTTA CCATTTATGTTTA CCATTTATGTTTA	650 ACTITCCGT ACTITCCGT ACTITCCGT ACTITCCGT ACTITCCGT ACTITCCGT ACTITCCGT ACTITCCGT ACTITCCGT ACTITCCGT ACTITCCGT ACTITCCGT	
200		\sim	nfi	dor		20				Traco	DOS	itions sh	iftod			
Jas	d to	,U			iut oto	53 101	t vol			Hate	pus	110115 511	nieu			
e.g.	10)0	a C		ວເດ	ai i (l vai	ue,								
_										Trace	display	/				•
📕 Loc	k Coli	umn	is: 1 2	34	Roy	ws:	1234	5 6 🗆 Sh	ow confi	dence Sav	e setting	s 📕 🔳 Compact			Close	e Help >>
#+1167 X Y	?? p /	EIK	K 1 49E 0)2BD6	610						-					



GAP4 assembly <u>without</u> application of sff_dump.c and roche454ace2caf.pl



GAP4 assembly <u>with</u> application of sff_dump.c and roche454ace2caf.pl

In conclusion, application of the two programs significantly improves the GS20 data exploitation in GAP4 by visualization of the aligned traces as well as availability of real base confidence values and hidden data. Nevertheless, there are still open questions for visualisation, particularly due to non-documented deletion of bases during the Phoenix assembly.

[1] http://staden.sourceforge.net[2] http://www.sanger.ac.uk/Software/formats/CAF