

Improved visualisation of GS20 traces in GAP4

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

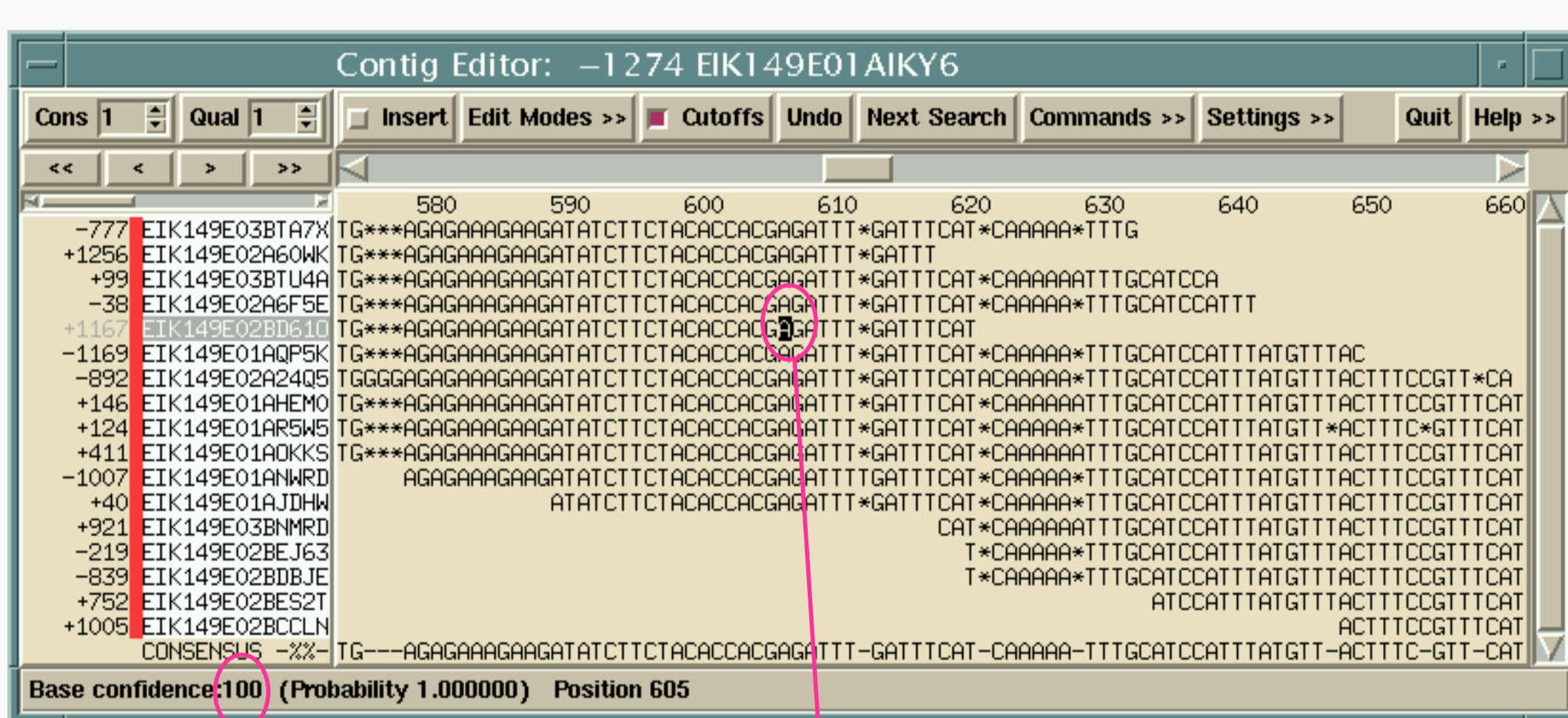
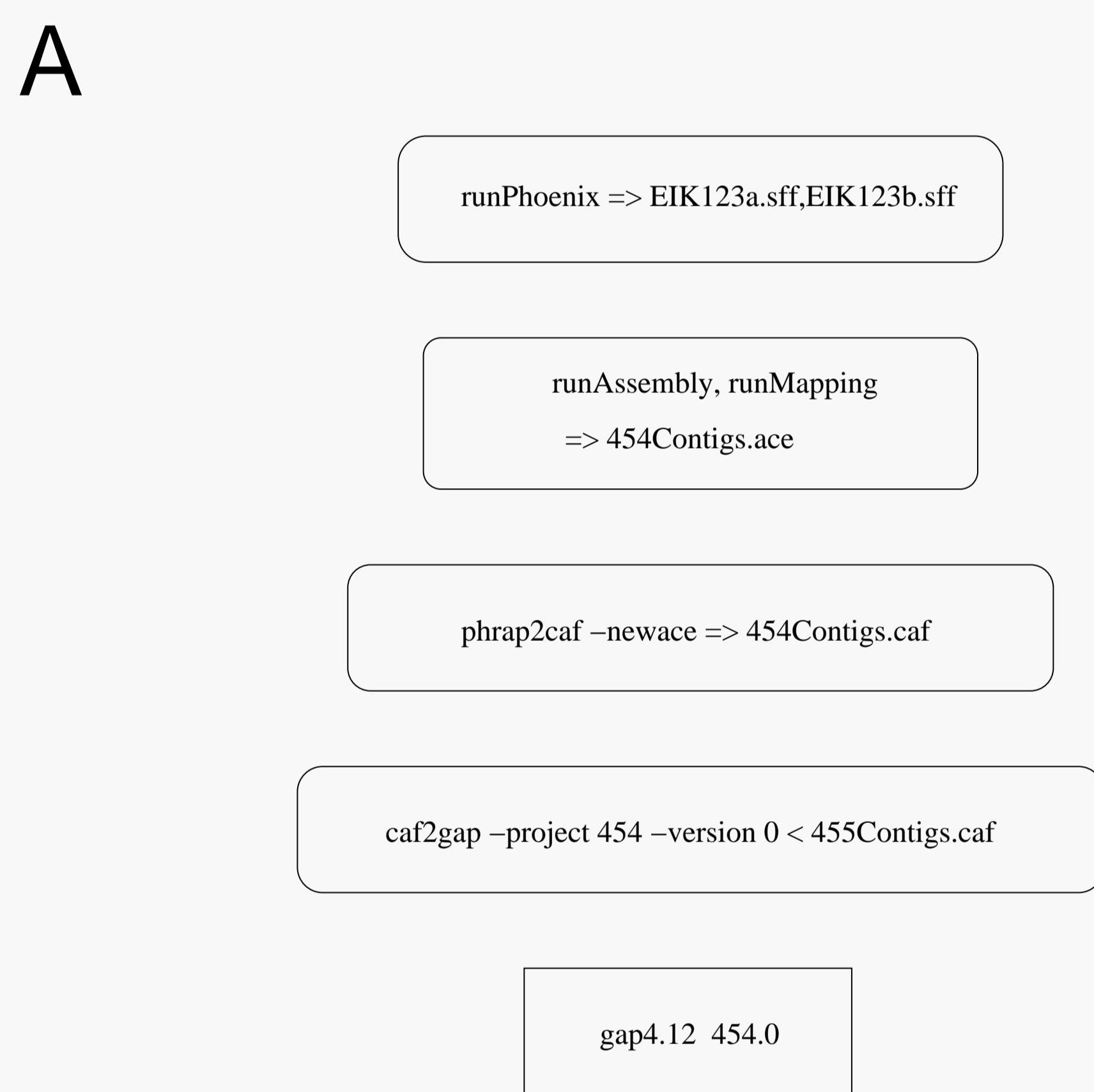
- a) Genome Analysis, Leibniz Institute for Age Research, Fritz Lipmann Institute (FLI)
Jena, Germany *) bsenf@fli-leibniz.de
b) Sanger Centre, Wellcome Trust Genome Campus, Hinxton
Cambridge, CB10 1SA, UK

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:

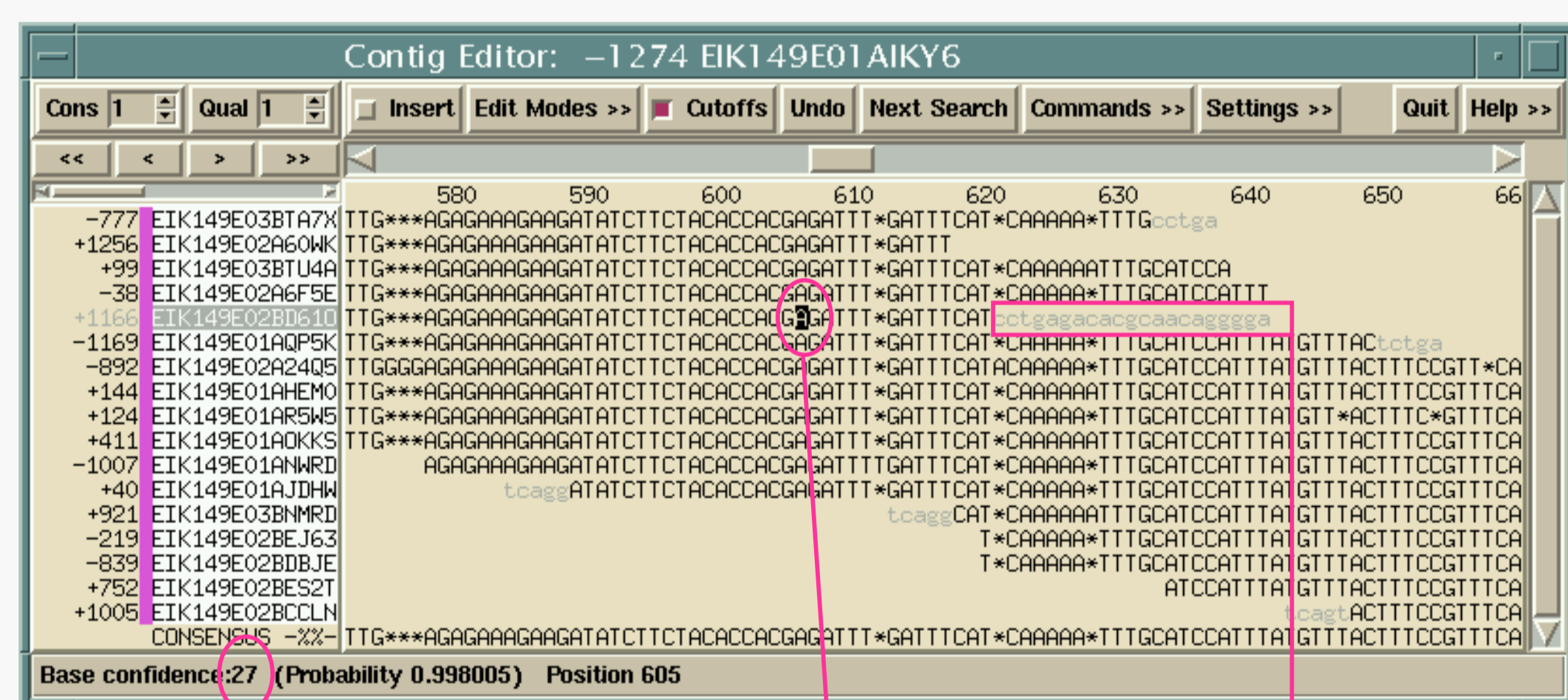
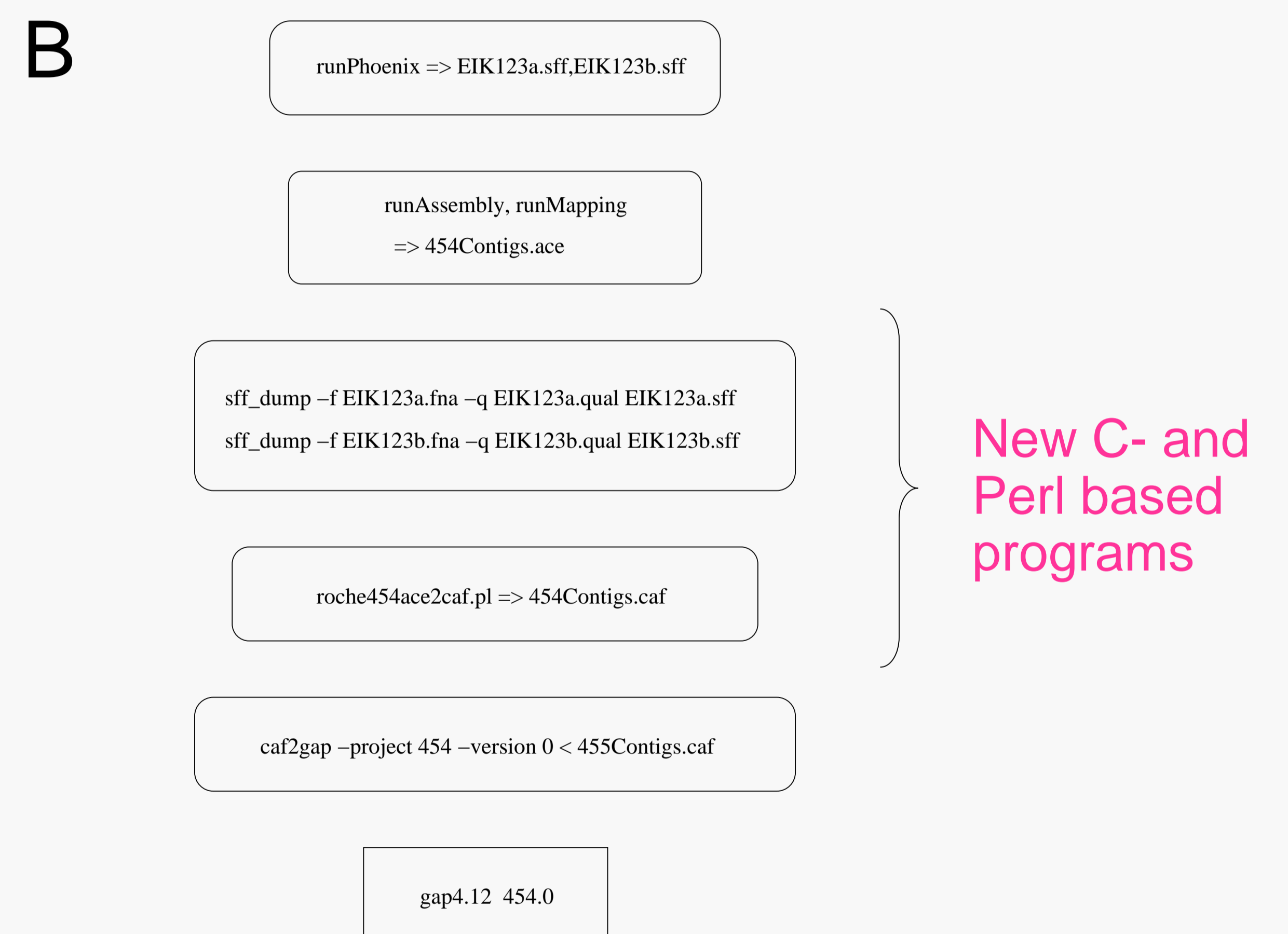
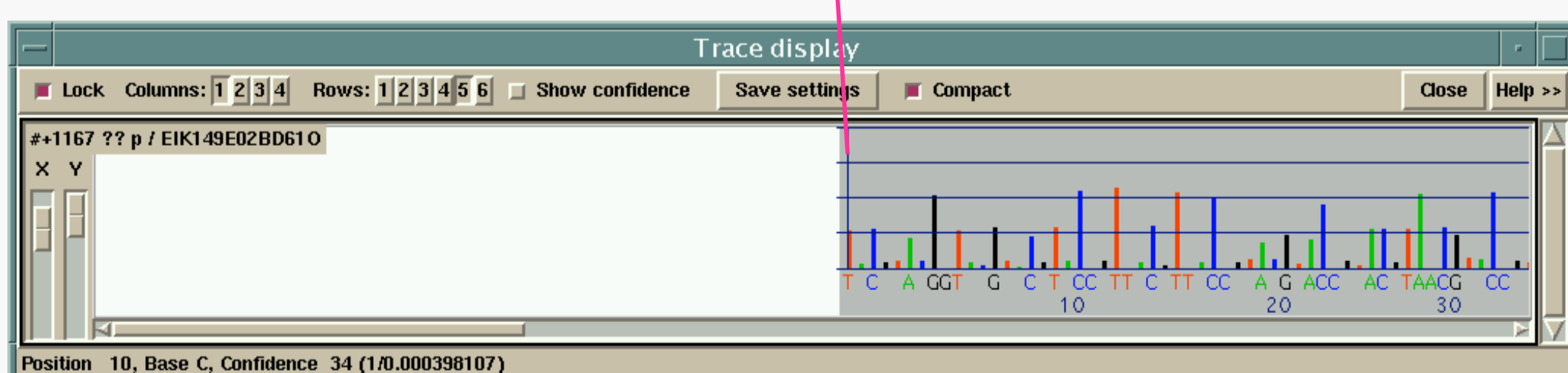
- Base confidence values are not available,
- Sequence parts that were excluded by the assembly program (“hidden data”) can not be retrieved,
- The pyrograms (“traces”) in their standard flowgram format (sff) can be visualized by a mouseclick, but the trace position is shifted with respect to the appropriate 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:

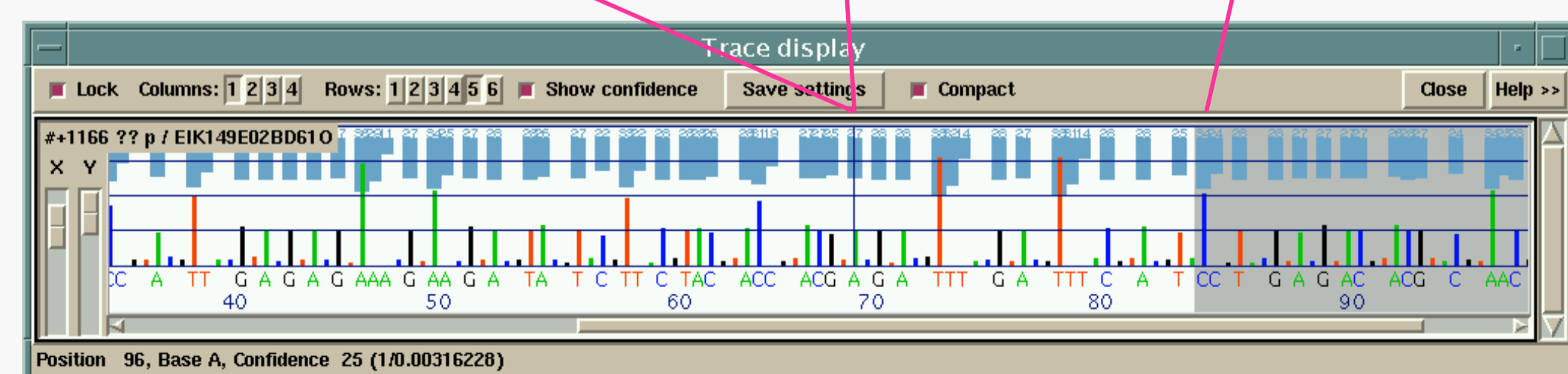
- visualizing the entire sequence, including “hidden data”, in the contig editor,
- assigning base confidence values to all read positions in the contig editor,
- visualizing the trace data at their correct positions



Base confidences fixed to a constant value, e.g. 100
Trace positions shifted



Real base confidence available
“Hidden data” available
Correct trace position



GAP4 assembly **without** application of sff_dump.c and roche454ace2caf.pl

GAP4 assembly **with** 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>