Title
Assembler Complementation Tool

Permalink
https://escholarship.org/uc/item/53g860j4

Authors
Singan, Vasanth
Trong, Stephan
Goltsman, Eugene
et al.

Publication Date
2006-05-03
A FEW UNDISPUTED CLAIMS: Misassemblies are likely to be caused in repeat regions. Misassemblies are characterized by partial matches to the subject (assuming subject is assembled correctly). The Assembler Complementation (AC) tool can be used to view and isolate regions that differ between two assemblies. The break points of misassembled regions can easily be located and can be used to fix them.

About me:

I am a Visualization tool solely created for the purpose of comparing two assembled and sequencing regions likely to be misassembled. I am created as an add on to the main program and display results in a visually intuitive manner. The tool is in beta stage and is still being improved. The main purpose I exist is to visually show regions of a genome that are likely to have been misassembled or broken because of repeats.

Figure 1 shows possible misassembly of two contigs by partial match to subject. The grey regions show mismatch.

Figure 2 shows possible misassembly of two contigs. The junction between grey and solid colors is the point of misassembly.

Figure 3 shows misassembly caused by a repeat. The region has hits from multiple contigs.

Figure 4 shows a repeat causing the assembler to fail in extending a contig. The contigs can be joined by pulling reads from repeat region and filling the gap.

Figure 5 shows a collapsed repeat resulting in 3 contigs. Visualizing them can help resolve misassembly easily.

Figure 6 shows the subject contig breaking at a repeat.

Figure 7 shows a gap caused probably due to low quality region or poor coverage.

Figure 8 is a good example that demonstrates fragmentation of assembly and can be used to compare assemblies.

Figure 9 shows a gap probably due to low overlap. Some assemblers are more stringent and may not extend contigs in low overlap / poor quality regions.

Legends:
- Black Background
- Yellow bar represents subject contig (against which other contigs are aligned)
- Other Solid colors represent query contigs (each contig is pre-assigned a specific color)
- Grey bar shows mismatch regions of any contig

Even More undisputed claims:

Low quality / coverage regions can result in gaps. This tool can be used to visualize such gaps and help in scaffolds.

Q1. Do, Yet another tool for finishing ?
A1: Yes ☺

Q2. So, this really works ?
A2: duh !!!

Q3. What is the input to this program ? Don’t tell me it uses some inane format which makes no sense!
A3: The program reads raw blast output. All you have to do is blast the two assemblies you want to compare and give the result as input to the AC tool.

Q4. Can this fix misassemblies ?
A4: NO ! But you can see where the assemblies differ.

Q5. Can I compare more than two assemblies at a time ?
A5: NO ! But you sure can use it over and over again, with different inputs !

Q6. Can I store the output ?
A6: You can either take screen shots of the contig window or you can click on a hit and the blast hit information is printed along with coordinates that can be copied and stored for future reference.

Q7. What feature are you planning to add for the next release ?
A7: Filing of query hits based on user selection and using alignment format other than Blast are the features we would like to have in the next release.

Q8. What programming language was used for this tool ?
A8: Perl with TK for graphics.

Q9. Are you thinking of renaming this tool ever? Don’t you think ASSEMBLER COMPLEMENTATION TOOL is a bit ?
A9. Aye I agree, it is a bit . Do you have any suggestion ?

On a happier note, “Finishing for dummies”, 2nd edition featuring the Assembler Complementation tool is coming to stores near you.