Assembler Complementation Tool

Vasanth Singan, Stephan Trong, Eugene Goltsman, Alla Lapidus

Joint Genome Institute – Production Genomics Facility, Walnut Creek, CA

JGI — DOE JOINT GENOME INSTITUTE — US DEPARTMENT OF ENERGY — OFFICE OF SCIENCE

About me:

I am a Visualization tool solely created for the purpose of comparing two assemblies and uncovering regions likely to be misassembled. I am a relatively new tool and I am proud to be one of the first and very few tools that strive towards using multiple assemblies from various assemblers to improve the quality of life of finishers who use me.

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.

I can reliably display gaps, and repeat regions. I've also been recommended for use with post-finishing analysis and comparing closely related genomes.

Do read the F.A.Qs and what my creators have to say about me.,

---

F.A.Qs

Q1. Oh, 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.

Q3. Can this fix misassemblies ?
A3. NO ! But you can see where the assemblies differ.

Q4. Can this tool, what is it called, yeah the Assembler Complementation tool, can it close gaps ?
A4. Nah ! But you can see if any assembler failed to join contigs while another assembler put them together.

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. Is there a way I can 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 co-ordinates that can be copied and stored for future reference.

Q7. What new features are you hoping to add for the next release ?
A7. The filtering of query hits based on user selection and using alignment formats 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. I have a real troublemaker genome. Can I try this tool now ?
A9. The tool is still in developmental phase and we are looking to formally release it to the public soon. Leave your contact info with one of the authors if interested.

Q10. Are you thinking of renaming this tool ever? Don't you think ASSEMBLER COMPLEMENTATION TOOL is a bit...?
A10. 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…

Finishing for DUMMIES — A Reference for the Rest of Us!
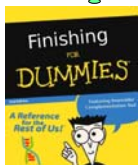
A FEW UNDISPUTED CLAIMS: Misassemblies are likely to be caused in repeat regions. Misassemblies are characterized by partial matches to the subject (assuming subject is assembled correctly). The Assembler Complementation (AC) tool can be used to view and isolate regions that differ between two assemblies. The break points of misassembled regions can easily be located and can be used to fix them.
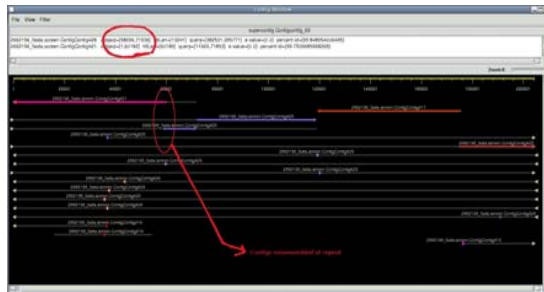

Figure 1. shows possible misassembly of two contigs by partial match to subject. The grey regions show mismatch.
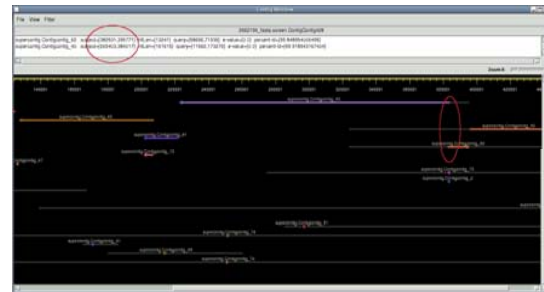
Figure 2. shows possible misassembly of two contigs. The junction between grey and solid colors is the point of misassembly.
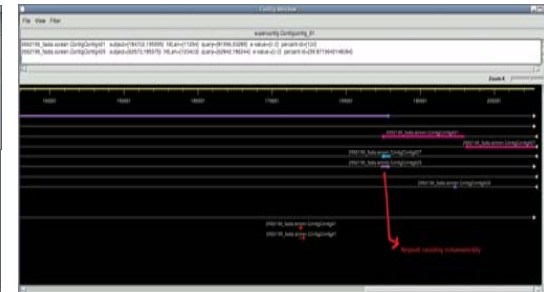
Figure 3. shows misassembly caused by a repeat. The region has hits from multiple contigs

---

SOME MORE UNDISPUTED CLAIMS: Repeats can prevents assemblers from extending contigs. Collapsed repeats are usually problematic and visualizing them can help us resolve misassemblies better. The screen shots below show specific cases where the assembler failed to join contigs because it could not resolve repeats.
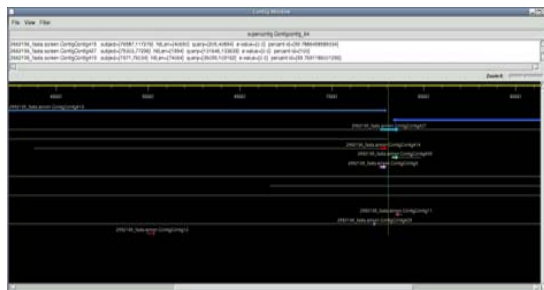

Figure 4. shows a repeat causing the assembler to fail in extending a contig. The contigs can be joined by pulling reads from repeat region and filling the gap.
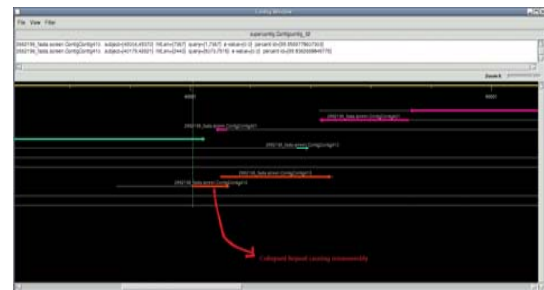
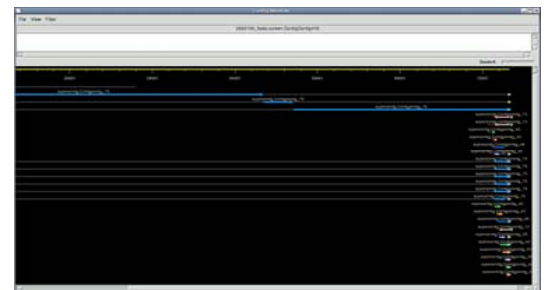Figure 5. shows a collapsed repeat resulting in 3 contigs. Visualizing them can help resolve misassembly easily.

Figure 6. shows the subject contig breaking at a repeat.

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.
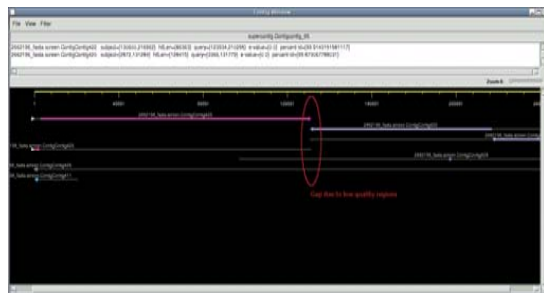

Figure 7. shows a gap caused possibly due to low quality region or poor coverage.
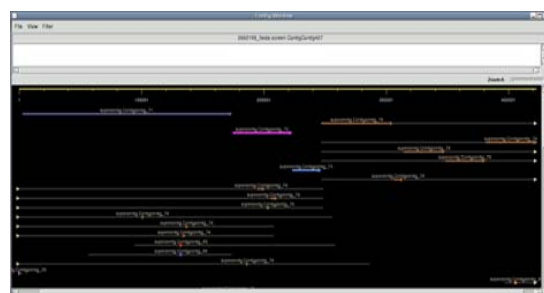
Figure 8. is a good example that demonstrate fragmentation of assembly and can be used to compare assemblies.
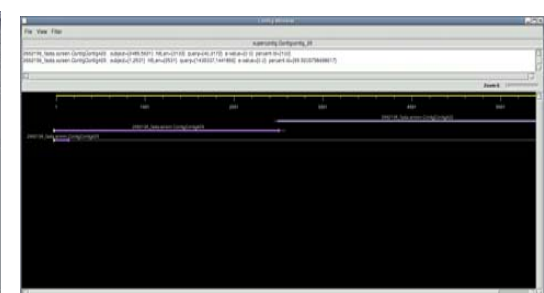
Figure 9. shows a gap probably due to low overlap. Some assemblers are more stringent and may not extend contigs in low coverage / poor quality regions.

* - All screen shots were taken using a blast of Arachne against Phrap draft assemblies of Dechloromonas aromatica, sequenced at JGI.

Legends

-- Black Background
-- Yellow bar represents subject contig (against with 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