

Lawrence Berkeley National Laboratory

Recent Work

Title

Genome Assembly Visualization Tool for Short Read Data

Permalink

<https://escholarship.org/uc/item/5dw3d2m0>

Authors

Aijazuddin, Syed
Trong, Stephan
Shapiro, Harris
[et al.](#)

Publication Date

2008-05-27

Abstract:

Genome Assembly Visualization Tool for Short Read Data

Aijazuddin Syed¹, Stephan Trong², Harris Shapiro¹, Eugene Goltsman¹ Kurt LaButti¹, Alla Lapidus¹ and Anthony Kosky¹

US DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598

Usually it is a challenging job for genome analysts to accurately debug, troubleshoot, and validate genome assembly results, including identifying such problems as mis-assemblies, low-quality regions, and repeats. Genome analysts rely on visualization tools to help with these tasks. Short read data generated by the new generation of high-throughput sequencing technologies add further complexity and make it extremely challenging for the visualization tools to scale and to view all needed assembly information. As a result, there is a need for new visualization tools that can scale to display assembly data from the new sequencing technologies.

We present GAViT, an assembly visualization tool developed at the DOE Joint Genome Institute (JGI), which can support data from new sequencing technologies and addresses the aforementioned concerns. GAViT incorporates numerous data handling, organization, and display optimizations, which make it interactive and scalable. In GAViT data is accessible at various levels of resolution, including scaffold, contig, read, and consensus. GAViT displays bird's eye views of assembly, inconsistently placed reads in the assembly, clone/read depth and GC graphs, quality scores, annotations, and read pair information by read or library type. A link viewer in GAViT provides a graphical view of how the contigs in a scaffold are arranged, and the linking reads within contigs are colored and oriented by the read type. Also, in GAViT we can edit library information on fly, group reads by type/library/insert/sequencing technology, generate various assembly analysis reports, and it has the facilities to add user annotations and user specific color schemes, and to view multiple contigs in either strand.

We have completed development and are currently testing GAViT with short-read data sets. The initial tests with traditional whole genome shotgun assembly data are very promising, scaling to view assemblies with 1.8 GB of genome sequence, and indicate that GAViT will be a scalable tool for visualizing large genome assembly data.

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.

¹ Lawrence Berkeley National Laboratory

² Lawrence Livermore National Laboratory