

Bacterial Genome Finishing Using Optical Mapping

Dibyendu Kumar, Fahong Yu and William Farmerie
 Interdisciplinary Center for Biotechnology Research,
 University of Florida, Gainesville, FL-32611

Abstract

The cost-efficiency of modern next-generation DNA sequencing technology allows investigators to undertake genome projects that were not affordable earlier. These sequencing techniques also bring new problems in genome assembly and finishing. Our core laboratory has several ongoing bacterial genome projects addressing a variety of challenges to genome assembly and closure. Several factors contribute to these challenges; including sequence repeats versus read length, intrinsic sequencing errors, and genome rearrangements. Together these factors complicate genome closure when using shotgun DNA sequencing data alone. In the absence of a physical map, we adopted **whole-genome optical mapping** as a tool to validate bacterial genome assemblies. **OpGen, Inc.** (Gaithersburg, Maryland) prepared the optical maps used in these projects. Briefly, an optical map is a complete genome restriction map deduced from a number of partial restriction maps. Optical maps are generated by spreading carefully extracted genomic DNA onto a treated glass surface containing many narrow channels, followed by digestion in situ with restriction enzymes. About 50–100 overlapping partial optical contigs are combined by alignment software to produce a contiguous whole genome restriction map. The contiguous optical map can be aligned and compared with the in silico restriction map of contigs obtained from whole-genome assembly. We successfully used optical mapping for guiding the closure of four closely related bacterial genomes. The optical map not only orient scaffolds but also allowed us to identify assembly errors, which was not possible using shotgun DNA sequencing data alone. Thus, we conclude that, in order to ensure the accuracy of a finished bacterial genome and to accelerate overall finishing process, optical mapping is an important tool to de-novo assemblies generated by next-generation DNA sequencing.

Optical Mapping Method



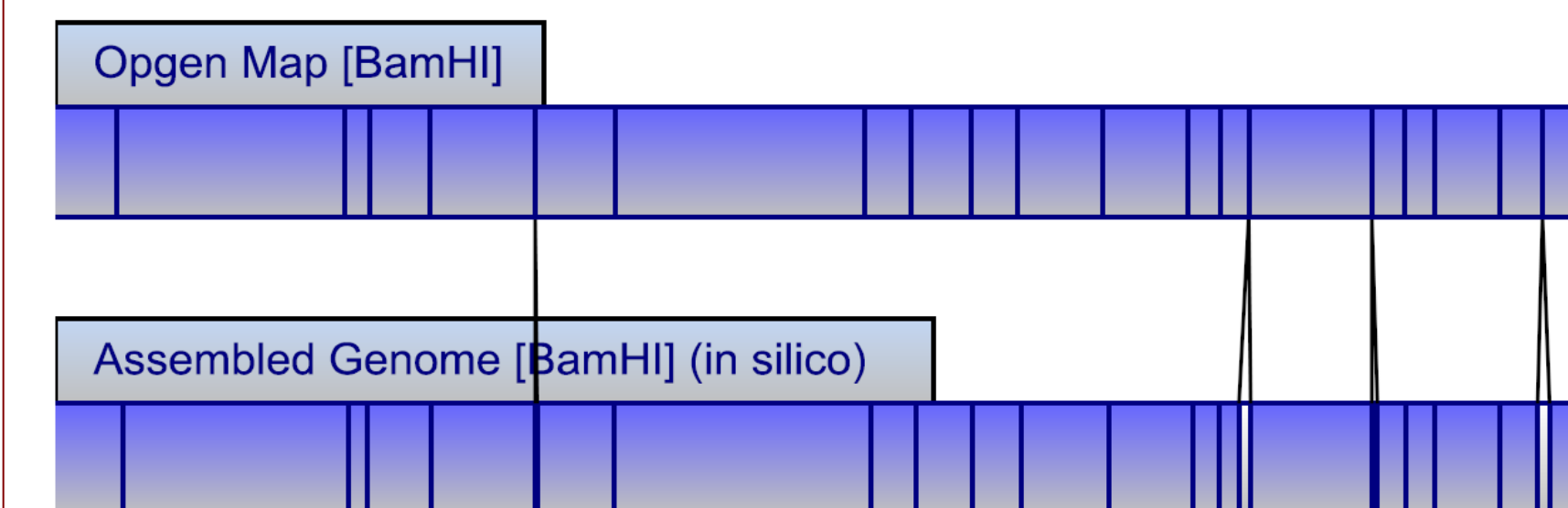
Optical chip containing single DNA molecule



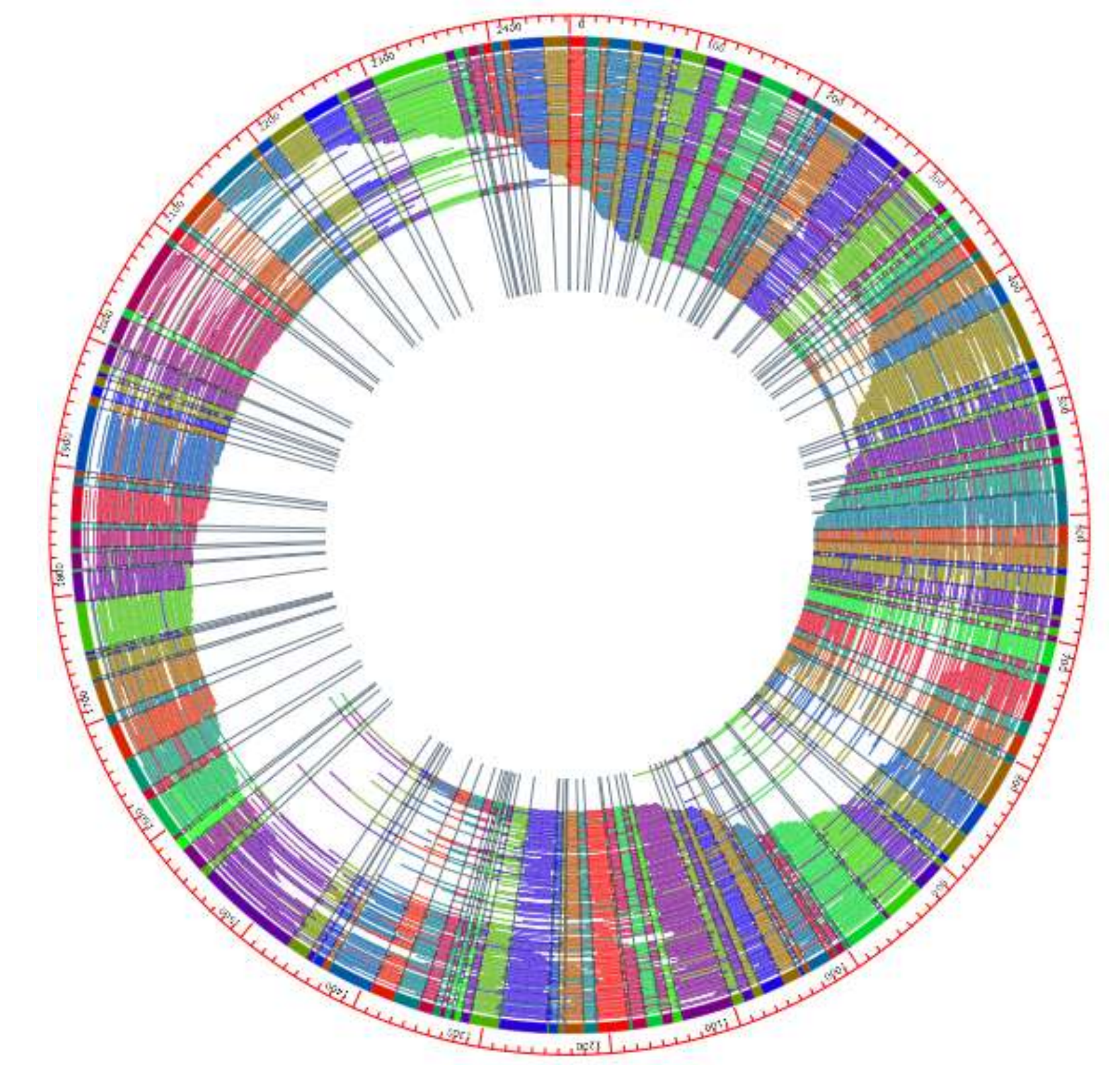
After digestion with restriction enzyme



Size and order of fragments deduction



Vertical lines indicate the location of restriction enzyme

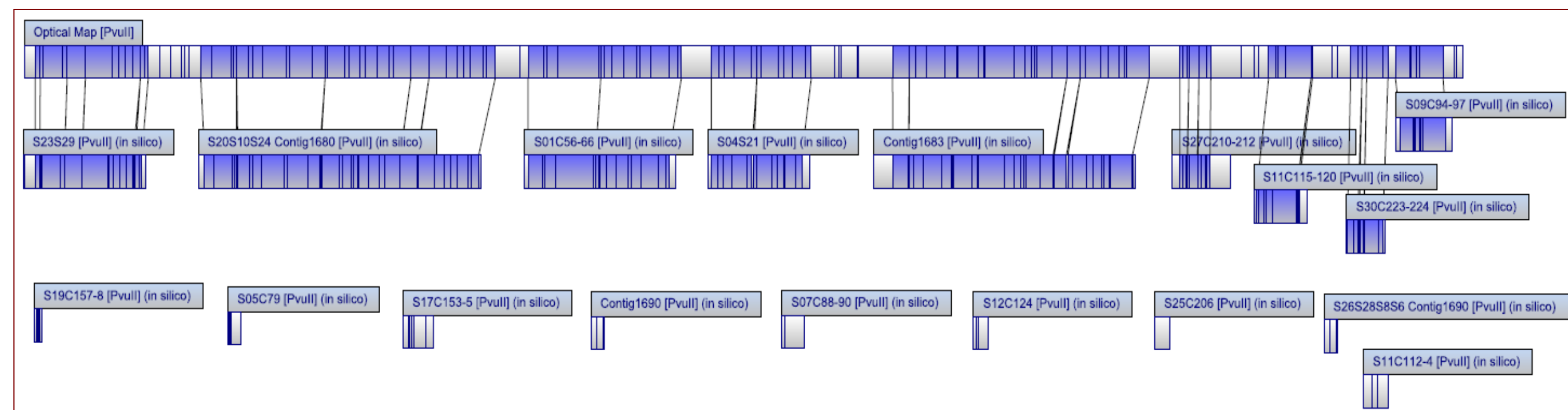


Optical circular *XhoI* map of a bacterial genome. The outermost red circle represents the consensus map created from the single-molecule maps shown as arcs. Different arc color is random and for contrast.

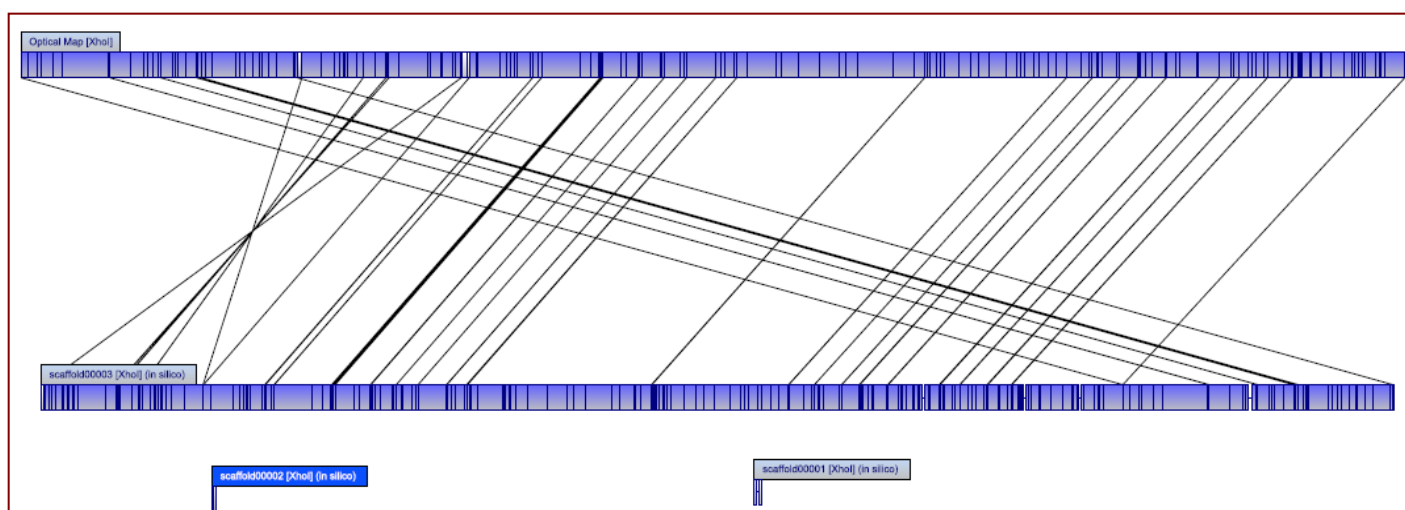
Challenges to Assembly

- No physical library, unknown Gap sizes
- Sequence read length
- sequencing errors such as, carry forward, homopolymer length, incomplete extension, etc
- genome rearrangement

Anchoring and Orienting Contigs to Optical Map



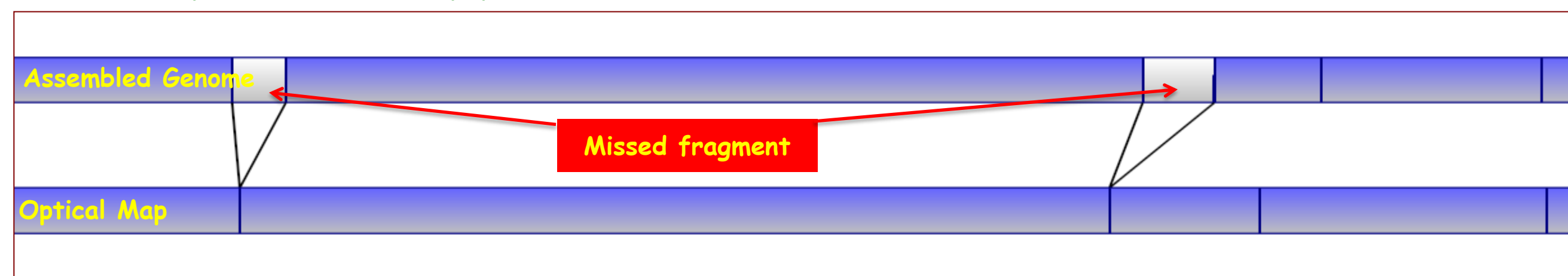
Detecting Misassembly



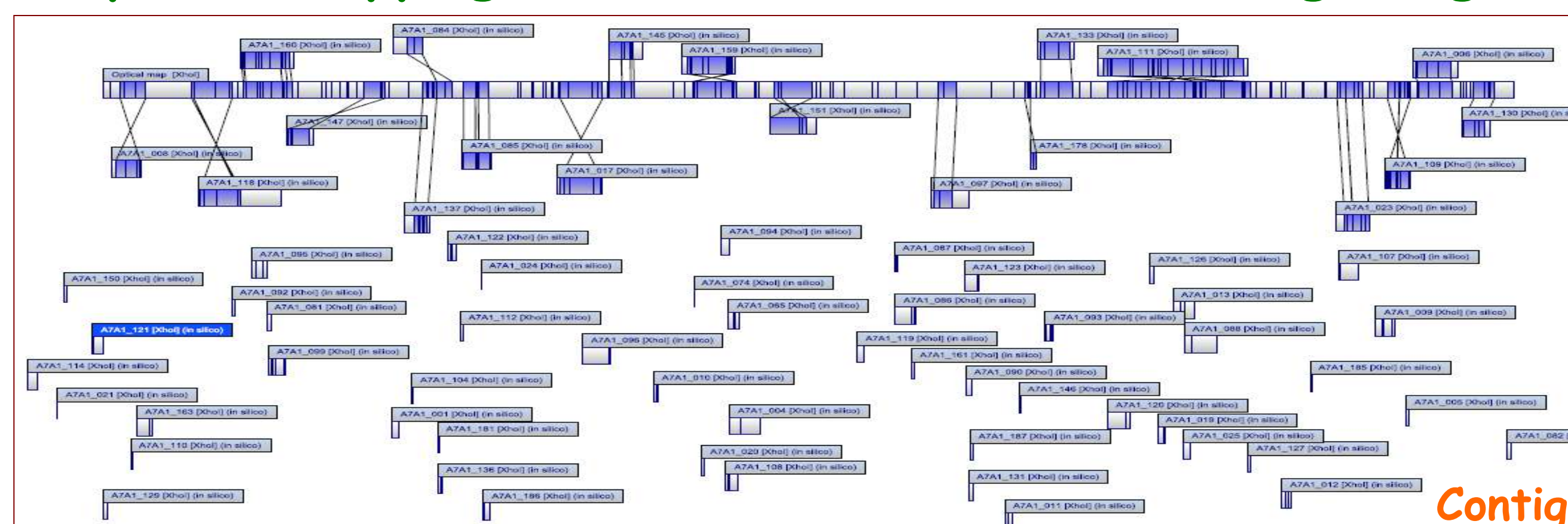
Detecting Deletion in Assembled Sequence



Optical Mapping Limitation: Missing Fragments

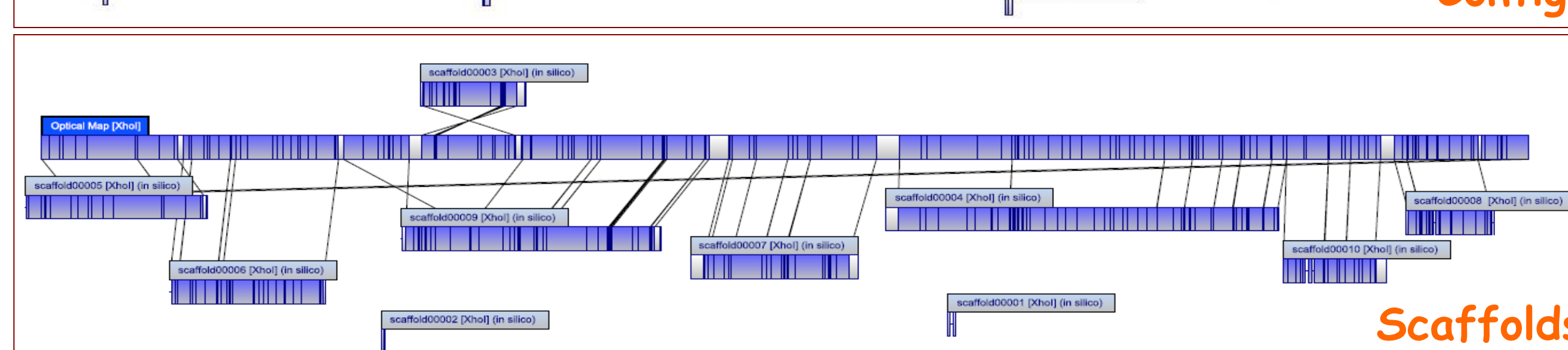
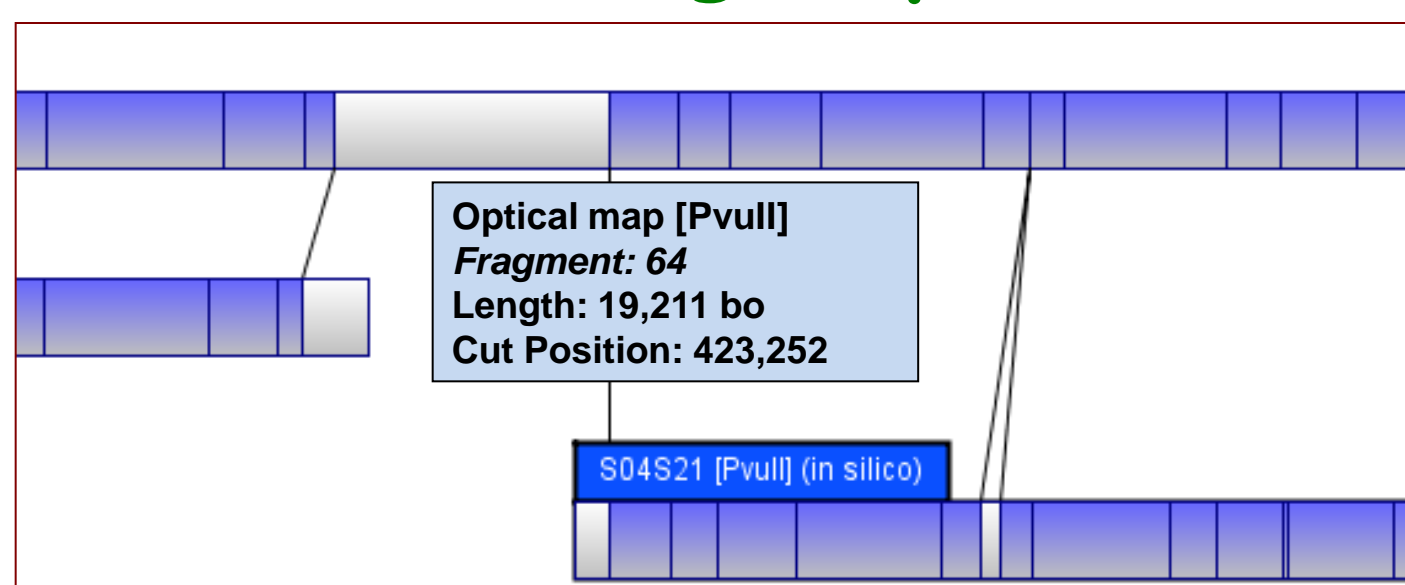


Optical Mapping Limitation: Assembled Contig Length



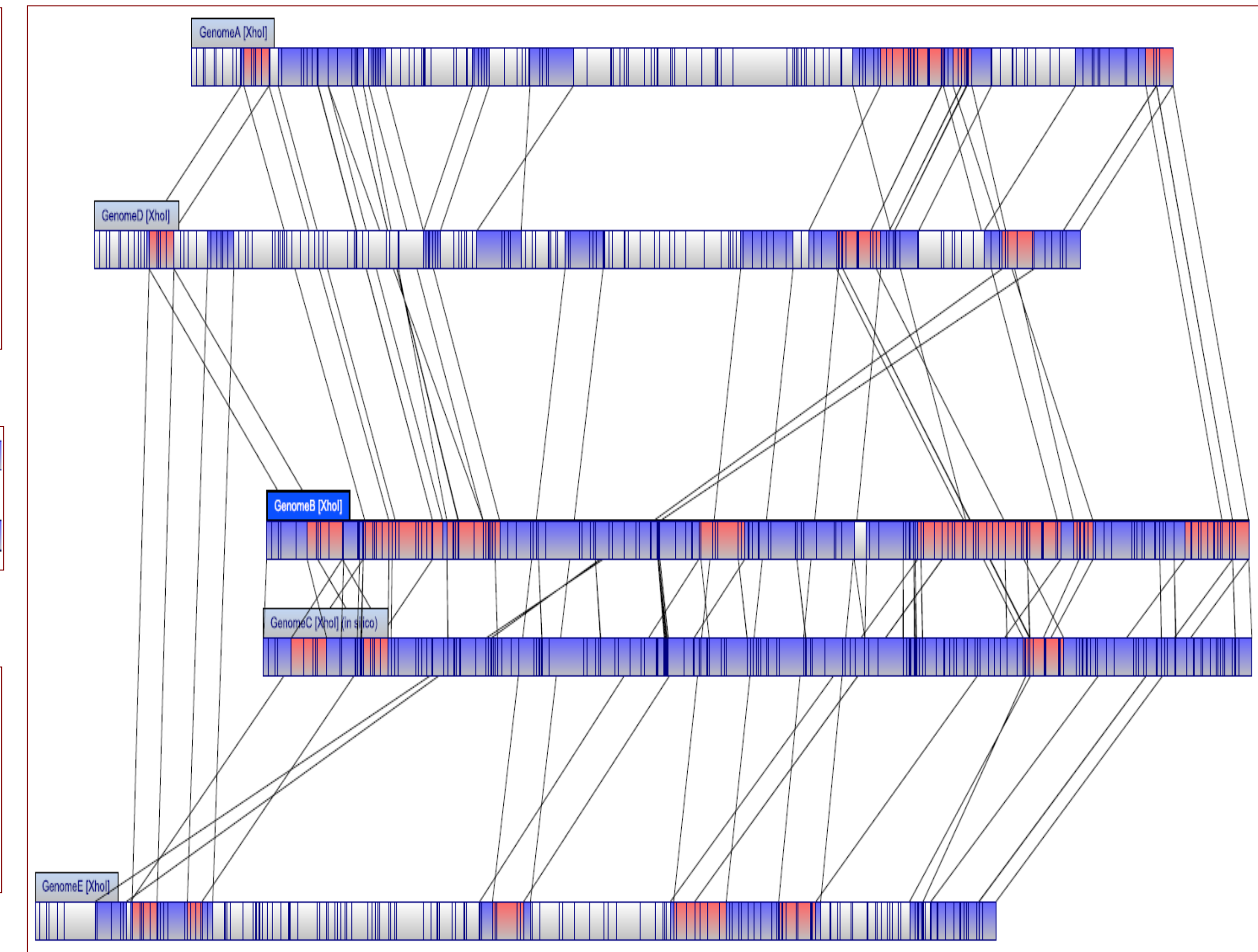
Contigs

Estimating Gap Size



Scaffolds

Optical Comparative Mapping



Conclusions

- ❖ Based on our experience, we strongly recommend including optical mapping in normal genome sequencing pipeline.
- ❖ It is relatively efficient, very fast and independent way to validate a bacterial genome assembly.
- ❖ Paired end reads are critical in building scaffold that can be aligned to optical map. Without paired end reads only a minority of contigs align to map. Many contigs remains as orphan.
- ❖ Effectiveness of optical map depends on choice of enzyme used for mapping. Sometimes, with some finishing jobs second mapping is critical.
- ❖ An optical map increases the speed of finishing and decreases the overall cost of the genome sequencing project.