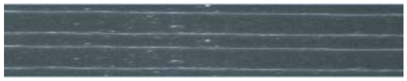
Use of Optical Mapping in Bacterial Genome Finishing

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

Abstract

The cost-efficiency of modern DNA sequencing technology, such as the Roche 454 GS-FLX, allows individual investigators to undertake bacterial genome projects that were not affordable only a few years ago. Our core laboratory has several ongoing bacterial genome projects presenting 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 dynamic genome rearrangements. Together these factors complicate genome closure when using shotgun DNA sequencing alone. The genome finisher may experience difficulty validating their assembly in the absence of a physical map. To address this problem, we adopted whole-genome optical mapping as a tool to validate bacterial genome assemblies. OpGen, Inc. (Gaithersburg, Maryland) prepared the optical maps used in this project. 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 contiguous restriction fragments with sizes approaching up to one-third of the whole genome are selected and optically measured. The 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 determined for the partially complete whole-genome assembly. We successfully used optical mapping for guiding the closure of four closely related bacterial genomes. The optical map allowed us to identify assembly errors not possible using shotgun DNA sequencing data alone. Thus, we conclude that, in order to ensure the accuracy of a finished bacterial genome, optical mapping is an important tool to validate *de-novo* assemblies generated by next-generation DNA sequencing.

Optical Mapping Method

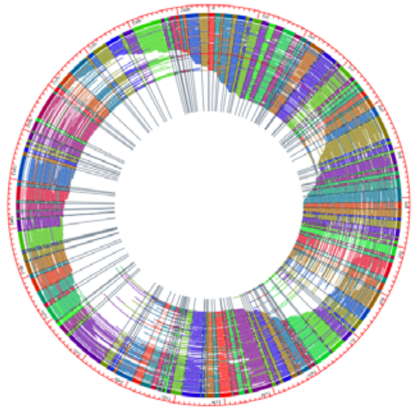


Optical chip containing single DNA molecule

After digestion with restriction enzyme

Software records the size and order of fragments

Optical Map, vertical lines indicate the location of restriction enzyme

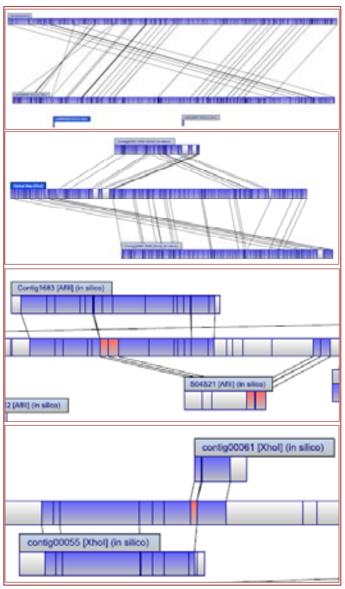


Optical circular XhoI map of *P gingivalis*. The outermost red circle represents the consensus map created from the single-molecule maps shown as arcs. Different arc color is random and for contrast.

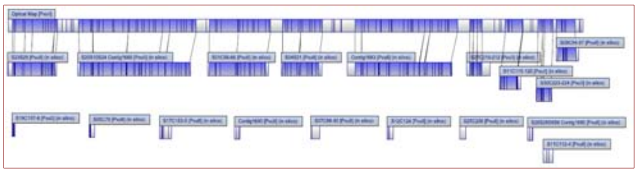
Challenges to Assembly

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

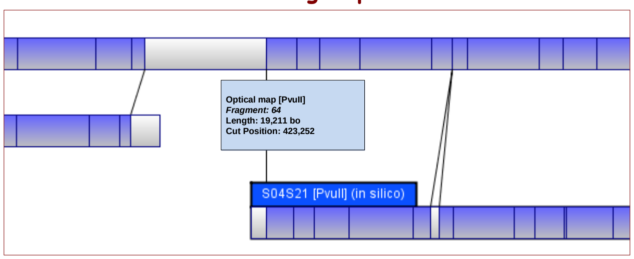
Detecting Misassembly



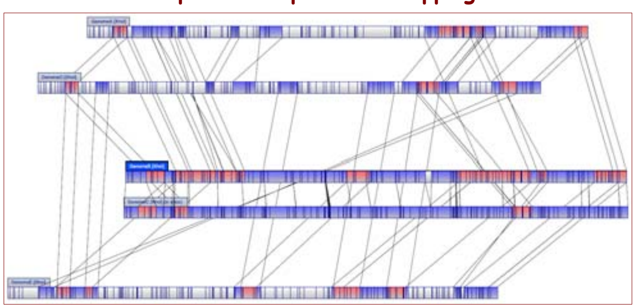
Anchoring Contigs to Map



Estimating Gap Size



Optical Comparative Mapping



Contig vs Scaffold Mapping

Conclusions

- Based on our experience, we strongly recommend including optical mapping in normal genome sequencing pipeline.
- It is relatively efficient 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.