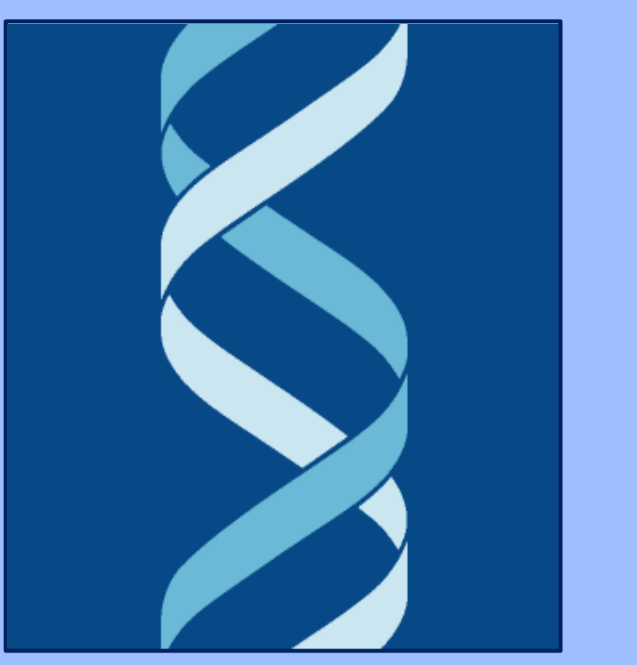


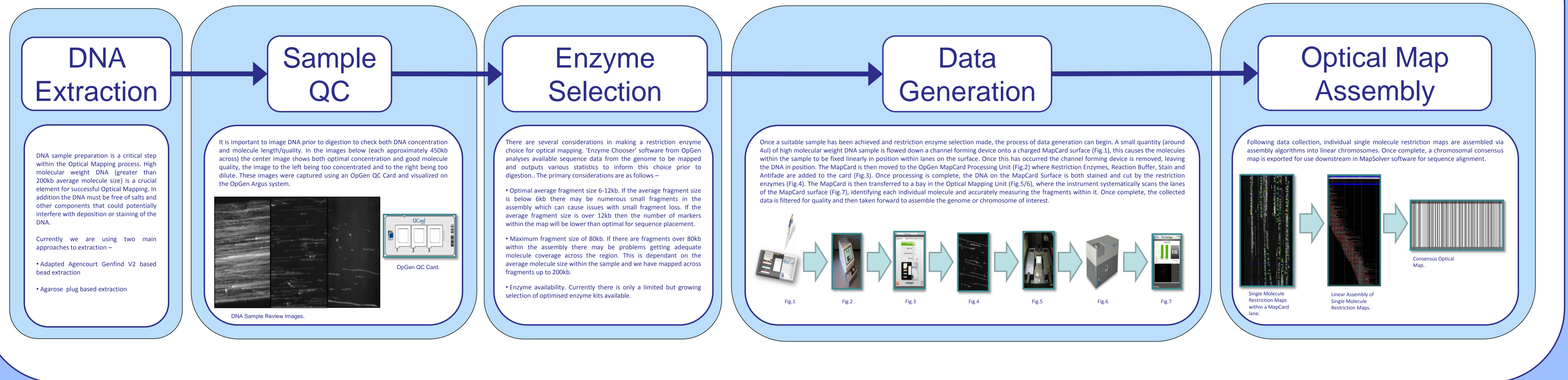
# Optical Mapping as a Complimentary Technology to NGS



Lucy Matthews and Matthew Dunn Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, CB10 1SA, UK

Optical Mapping is the process which allows the creation of a genome or chromosome sized restriction enzyme map of an organism, from very small quantities of high molecular weight DNA. The DNA is run through nanochannels, fixed in place, stained, digested and visualised using an optical microscope. The individual fragments within the molecules of DNA are then measured and the molecules are assembled together according to matching patterns of cleavage, thus creating a de novo restriction enzyme map. This entire process can be carried out within a week, a similar speed to the generation of NGS sequence. Scaffold contigs of NGS data can be digested *in silico* and aligned to the optical map, allowing ordering, orientation and gap sizing information to be inferred.

## Optical Map Production Pipeline



## Optical Map Utilisation

### MapSolver Software

MapSolver software from OpGen allows us to view, compare and align restriction maps. These restriction maps can be whole-genome Optical Maps produced via the lab process outlined or *in silico* maps that MapSolver creates from sequence data stored in FASTA or GenBank files.

With the MapSolver, we can:

- Perform genome comparisons between related organisms based on the restriction patterns of their restriction maps. The comparison enables you to identify insertions, deletions, duplications, and rearrangements.
- Identify motifs, annotate features and view sequence data from *in silico* maps.
- Perform sequence placement by aligning restriction maps of sequence contigs to an Optical Map for whole genome finishing and sequence validation. MapSolver enables you to validate assemblies, correctly orient contigs, and identify/design probes to close gaps and finish the assembly.
- Create similarity clusters for different strains and view the level of similarity. Clustering also allows the user to see the alignments that were used to calculate the cluster.

Sequence / DNA Alignment, H. meningitidis project

### Whole Genome Sequence Assembly

Unordered sequence contigs aligned to an Optical Map enables us rapidly to orient contigs and identify misassemblies.

Gap size ~ 700 bp (PCR to close)

Trim ends & join contigs

Expanded view showing contig alignment and gap measurement to facilitate closure and accelerate finishing.

### Sequence Assembly

Optical Mapping can be used to close gaps and validate whole genome sequence assembly both rapidly and cost effectively.

- Orient and align contigs to an Optical Map scaffold.
- Determine gap size and location.
- Find chromosomal inversions, insertions, deletions, and translocations.
- Identify and correct misassembled sequences.

### OM Project Example

#### C. jejuni Project

- Collaboration with Genome Science Centre, British Columbia Cancer Research Centre.
- *Cellulosa jejuni*, gram positive bacteria.
- 4.04 Mb of placed sequence (4.2Mb genome).
- Sample to sequence placement within 24 hours.

C. jejuni DNA Extraction QC Review.

C. jejuni OM Assembly.

C. jejuni OM/Sequence Alignment in MapSolver Software.

### Further OM Applications

#### OM Comparative Genomics

- Compare Optical Maps between related organisms and view differences and similarities.
- Find genetic variation: insertions, deletions and other genetic modifications.
- Track mobile elements, pathogenicity islands, phage, transposons.
- Distinguish strains & determine genetic relatedness between multiple isolates via Map-based clustering.

Multi-isolate comparisons reveal motifs associated with critical phenotypes such as *Mec* cassette variants conferring Methicillin resistance in *S. aureus*.

Directly compare maps of related isolates to locate novel differences, such as insertions (top panel), white segments or deletions (bottom panel). Regions of similarity indicated in green.

#### OM Strain Typing

- Perform high resolution epidemiology by Optical Mapping.
- Characterise and monitor strain stability.
- Distinguish strains and determine genetic relatedness between multiple isolates via Map-based clustering.

Map based similarity clustering precisely distinguishes strains and discerns relatedness between isolates.

### Argus System Overview

The OpGen Argus Optical Mapping system encompasses all the requirements to generate optical maps, including –

- Optical Mapper
- MapCard Processor
- Mapping Work Station
- Oil Applicator
- MapManager Software

Following map generation the system includes MapSolver software for downstream analysis, encompassing whole genome sequence assembly, comparative genomics and strain typing.

A range of complementary Argus optical mapping consumables are available including MapCard Kits, QC Card Kits, Sample Preparation Kits, Stain Kits and a range of Enzyme Kits.

## Future Modifications for the Optical Map Pipeline

- We are currently trialling a Large Genome Pipeline, which utilizes a merged assembly approach, incorporating both sequence data and optical map molecules. In this method, the pattern of restriction sites at the end of each sequence contig is used to identify overlapping optical map molecules; the optical map data joined at the end of the sequence contigs is then 'grown' until a join is found. Thus sequence data is brought together into scaffolds and sequence gaps sized. Testing is due to start shortly on a human cell line, with the aim of completing the process by October.
- We are working towards entirely de novo assembly of mid-range genomes of increasing size, with the aid of various lab process improvements, including new chemistry and higher density channel forming devices.

## References

- [www.opgen.com](http://www.opgen.com)
- Argus Optical Mapping Users Manual (OpGen)
- Argus Training Manual (OpGen)



## Acknowledgements

Thanks very much to all the people involved in the Optical Mapping Pipeline at the Sanger Institute and OpGen.

