

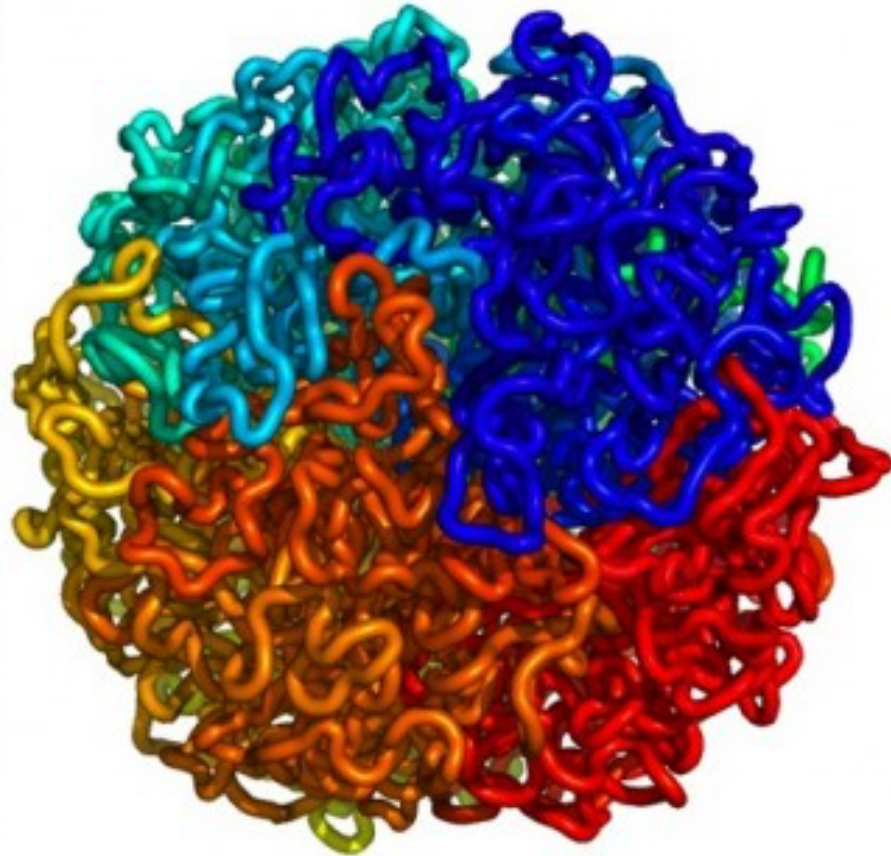
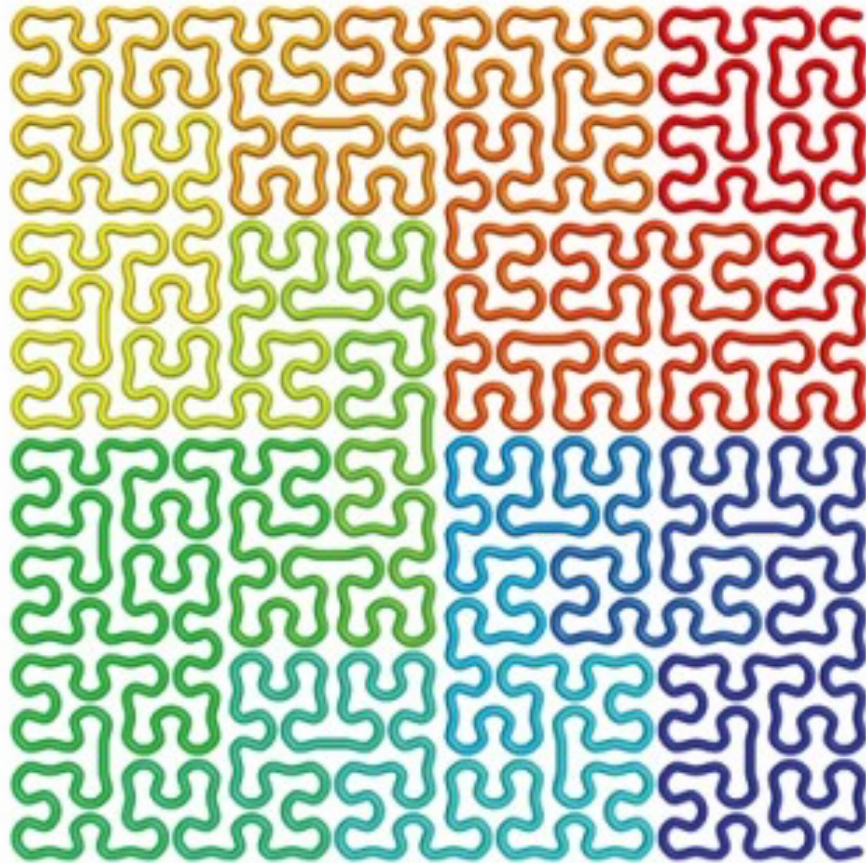
Disclosure*



PHASE
G E N O M I C S

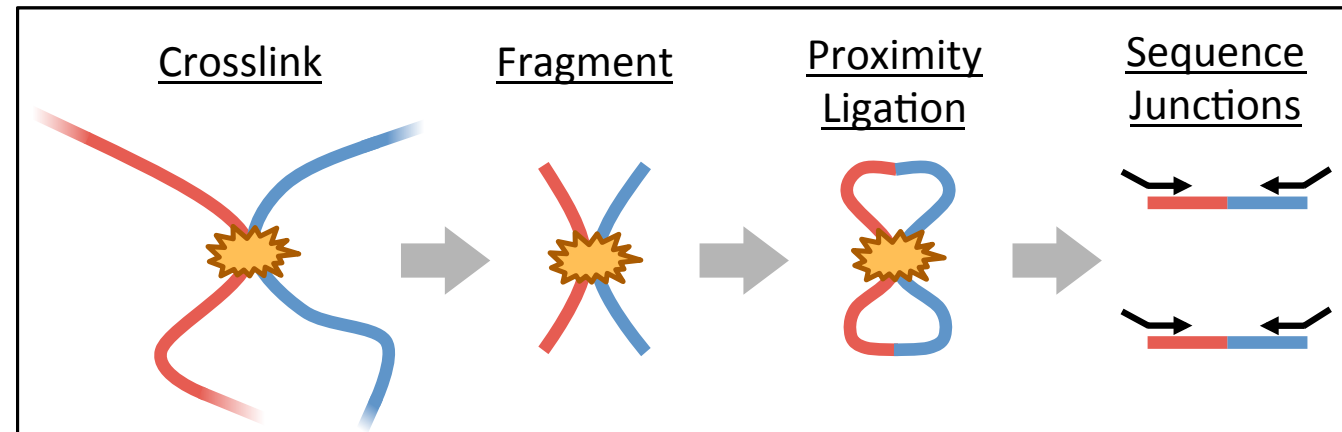
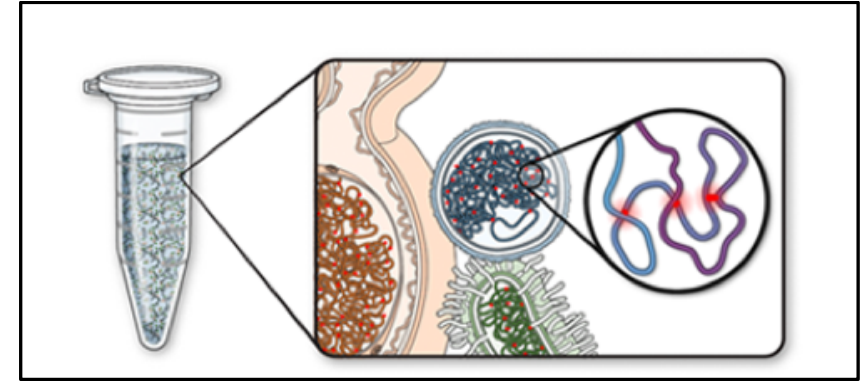
* We build high-quality PacBio genomes with Hi-C at Phase Genomics!

From contigs to chromosomes, Hi-C proximity-guided assembly



Hi-C measures the physical proximity of DNA

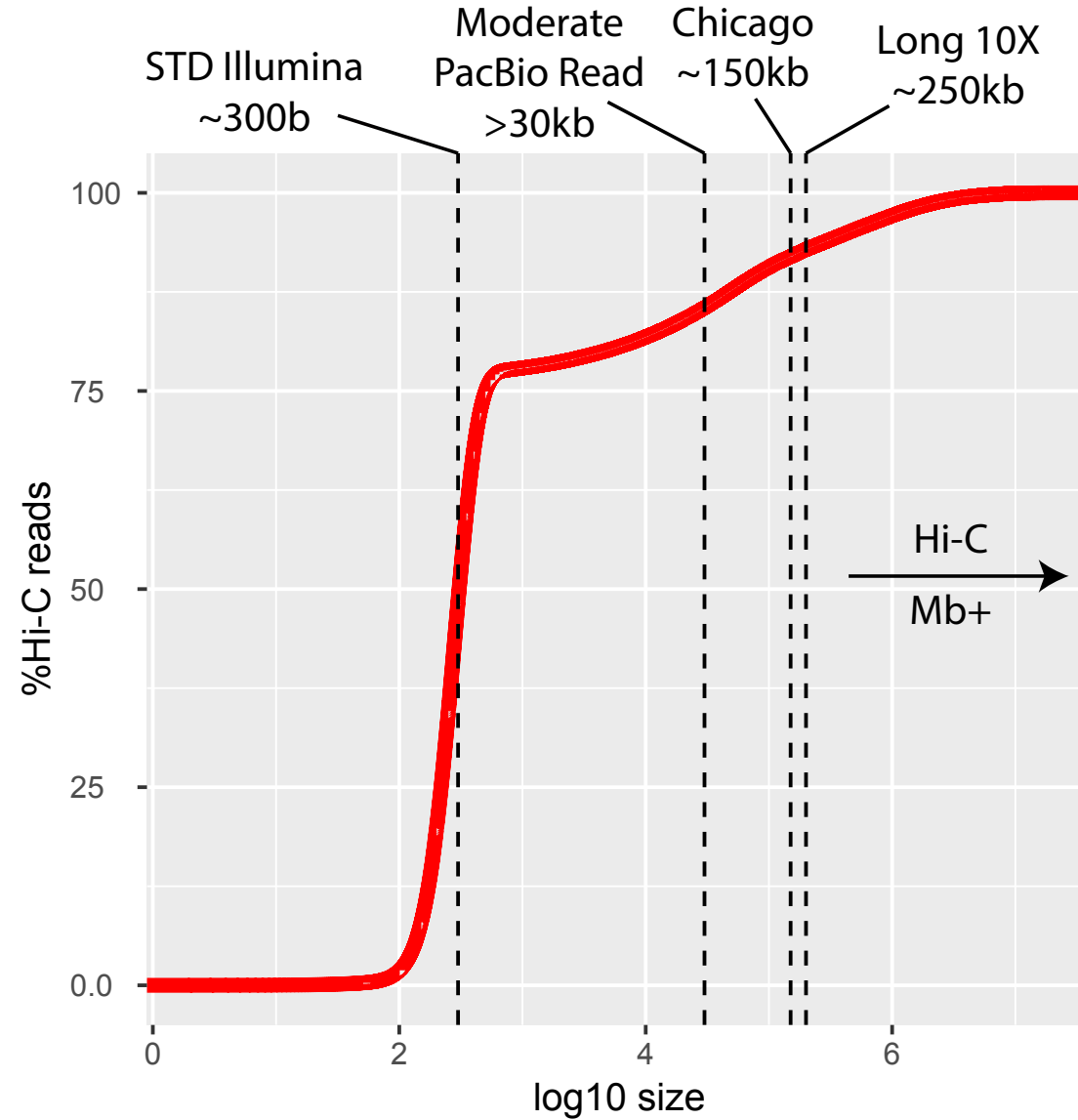
- Hi-C is a single cell technology
- The libraries are generated from cross-linked DNA
- The libraries enrich for chimeric molecules



- Std. PE library
- Compatible with all Illumina sequencers



Hi-C provides ultra long range information

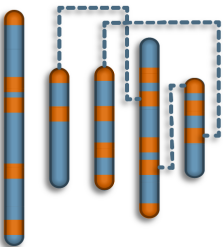
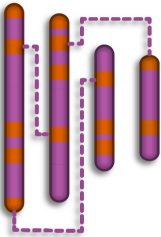
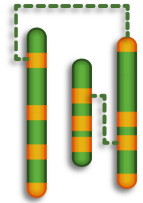


Hi-C data is useful for many applications

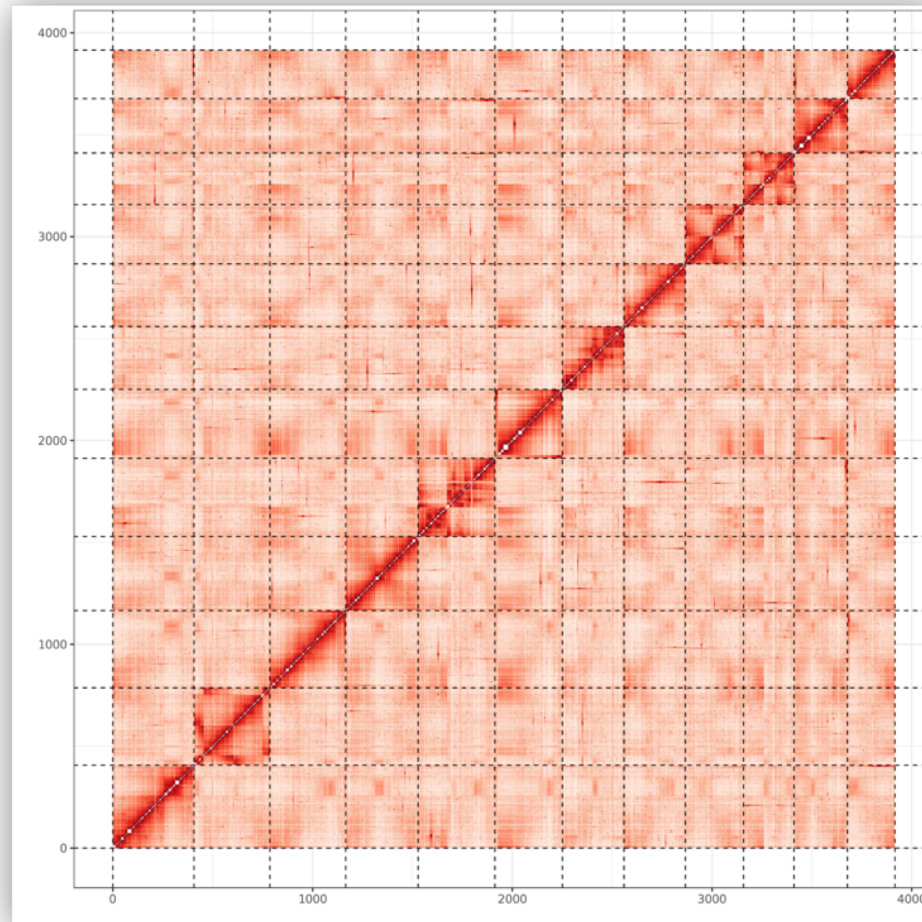
- Genome Scaffolding
- Metagenomics
- Genome Phasing
- Structural variation detection
- Epigenetics

Chromosome-level scaffolds achieved with Hi-C

- Hi-C data maps across contigs



- PacBio Contigs => Phase Genomics Proximo Scaffolds



• Hi-C Scaffolding tools

- Proximo
- SALSA
- 3D-DNA
- HiRise
- HiCembeler
- *et al.*

- Mint Genome, 2n=12, Dr. Kelly Vining O.S.U.

Several useful utilities for Hi-C scaffolding

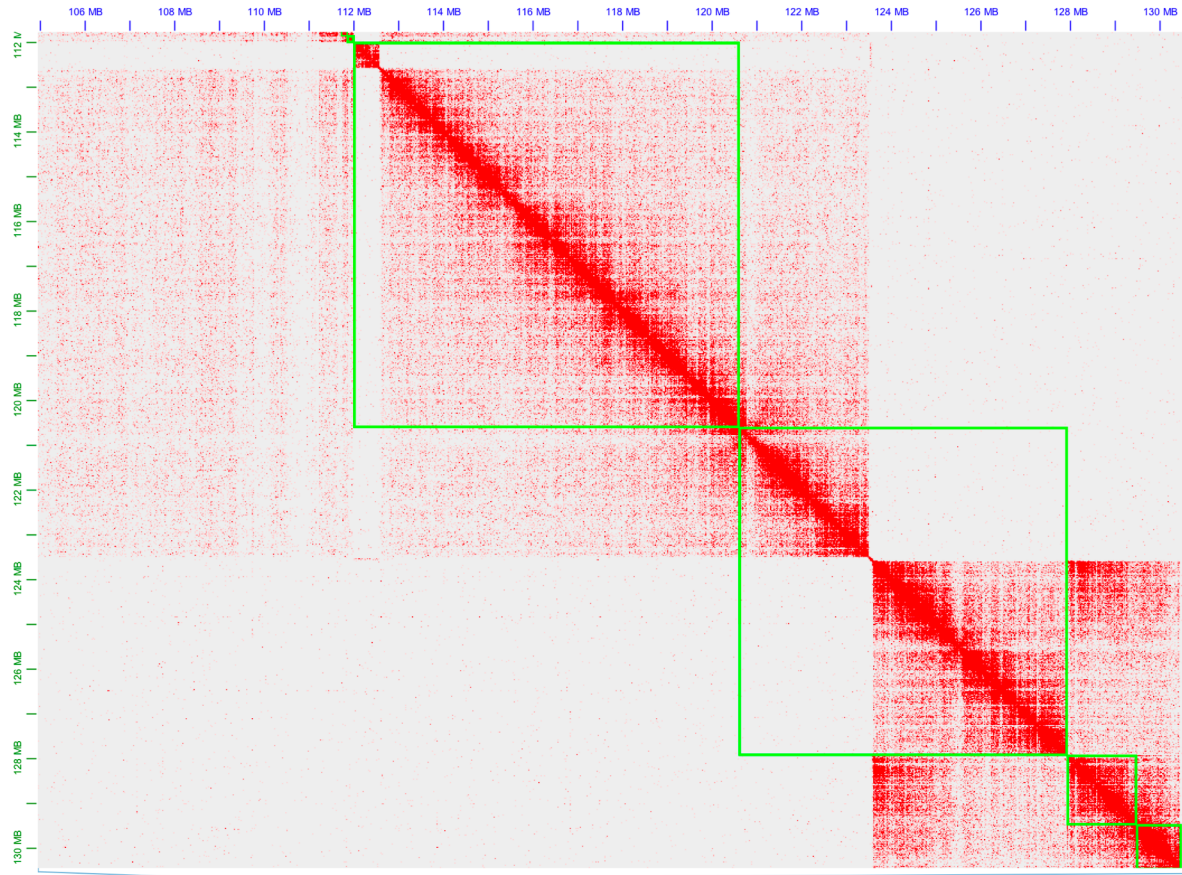


PHASE
GENOMICS

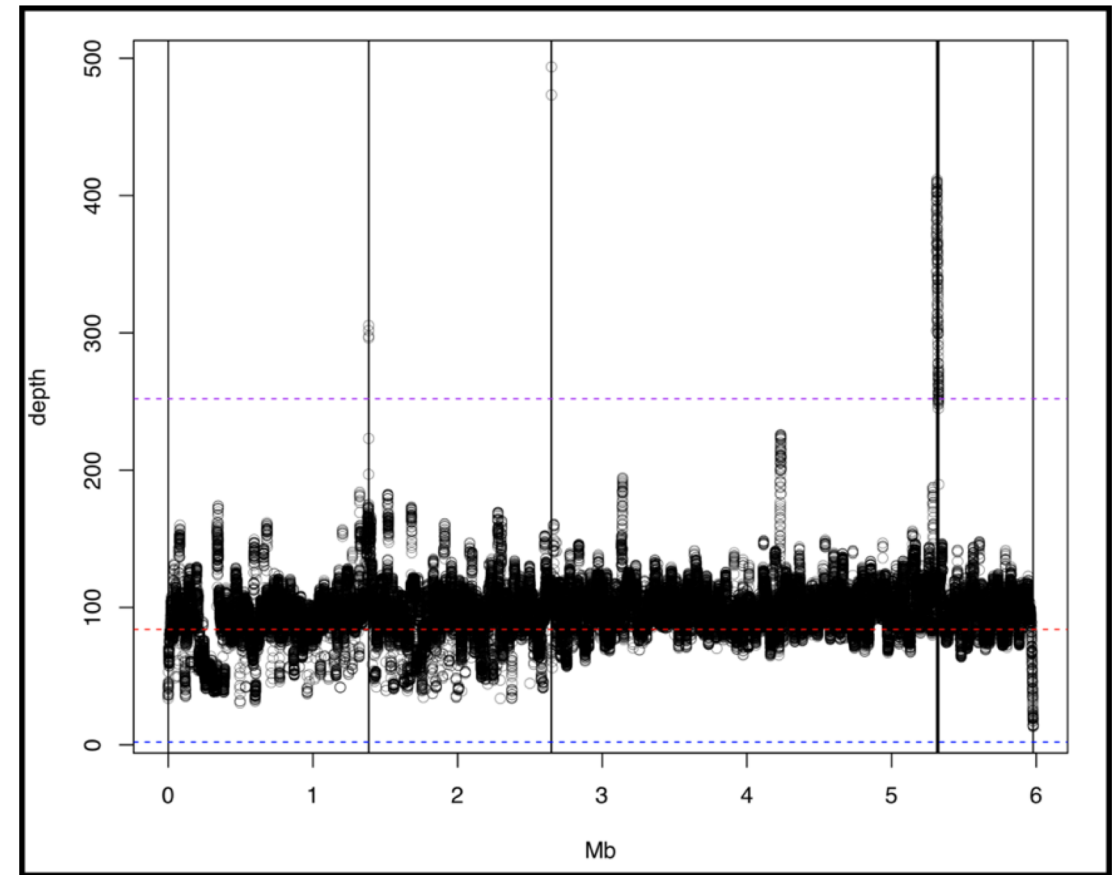
<https://github.com/phasegenomics>

- Polar Star
 - Break contigs on PacBio read depth
 - Simple heuristics
 - Reduces chimeric contigs
 - Minimal reduction in contig N50
 - Improves scaffolding
- Matlock
 - Advanced BAM/SAM/CRAM filtering
 - Inclusion and exclusion of contigs
 - Binning
 - Counting
 - Motif counting

Polar Star: contig breaking on long-read depth



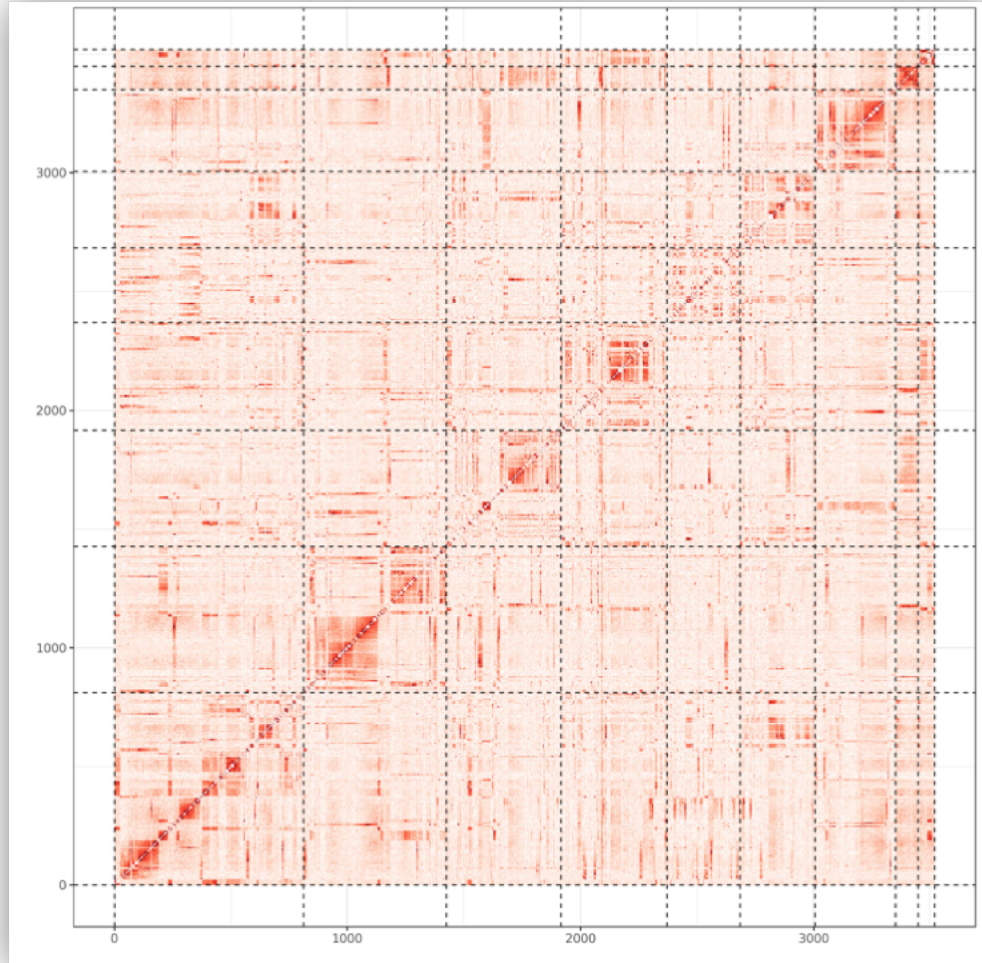
Two large-scale chimeric contigs



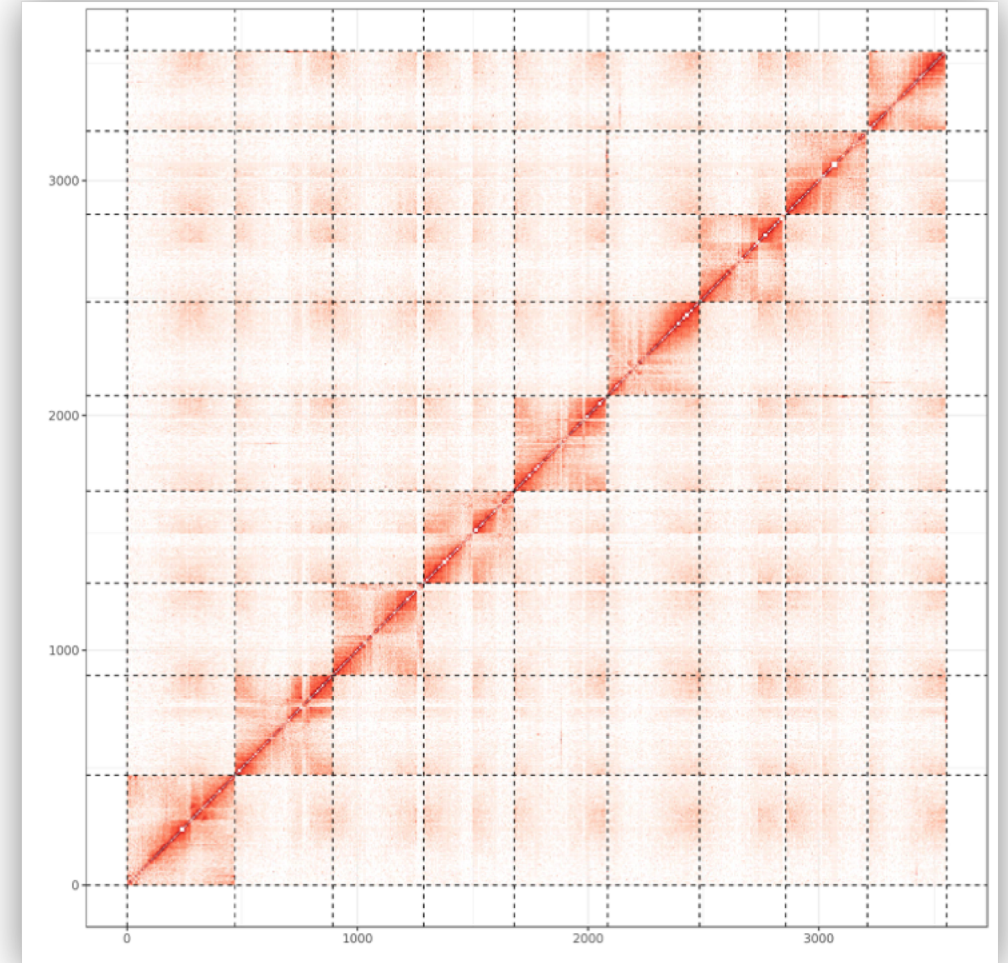
Depth profile along chimeric contig

Matlock filtering can improve clustering, ordering, and orientation

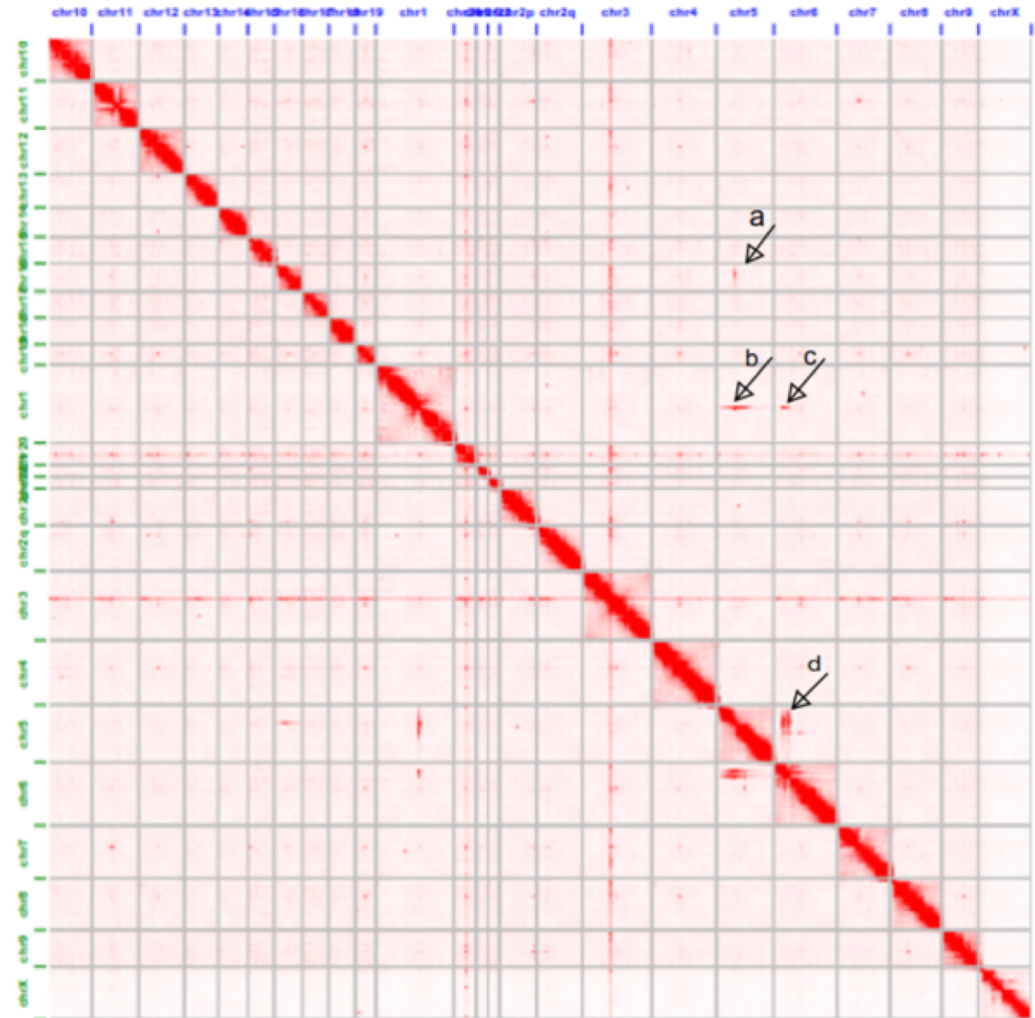
Unfiltered scaffolding (Kaniwa)



Post Matlock

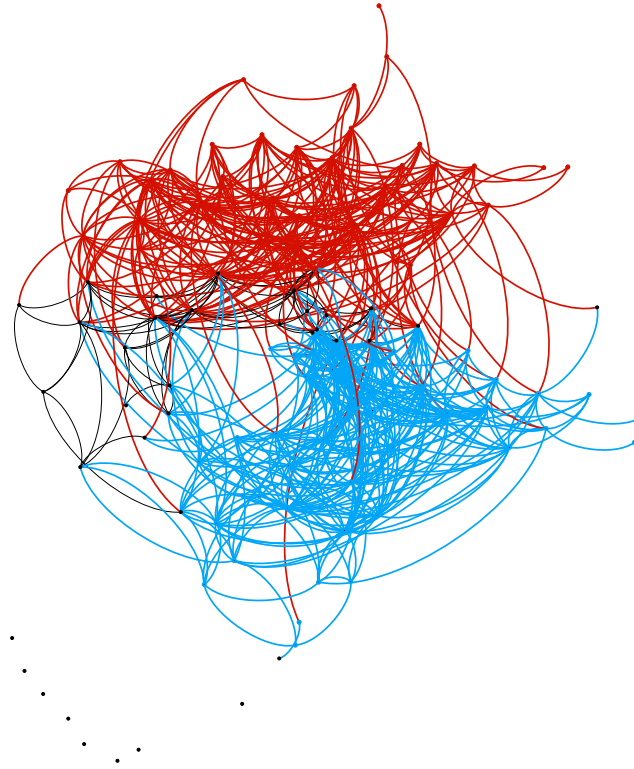
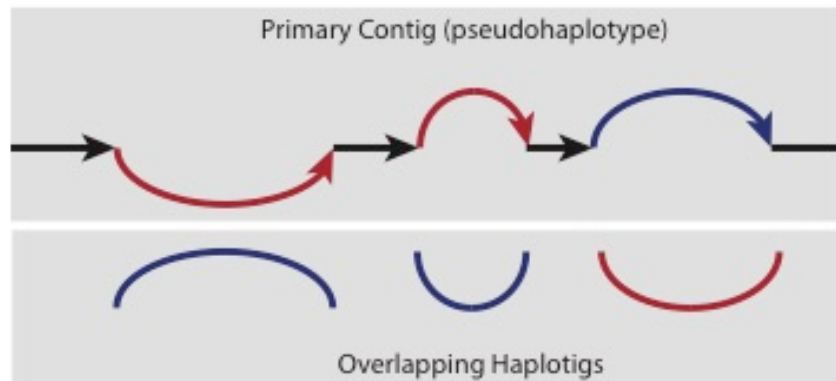
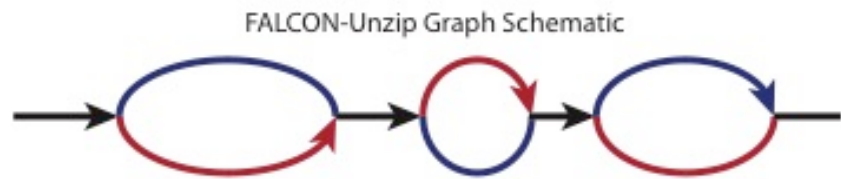
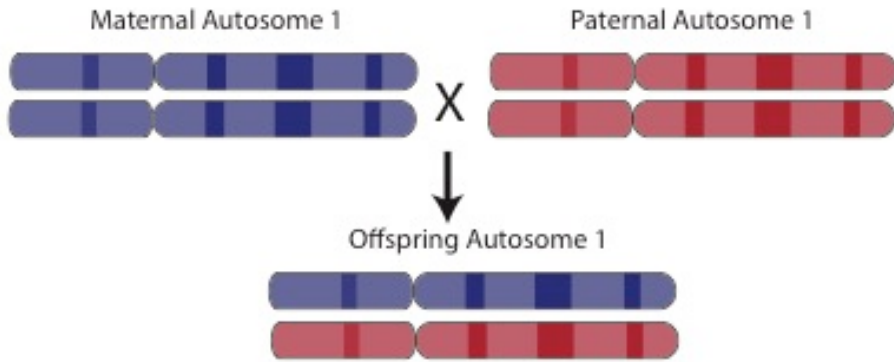


Hi-C for independent scaffold validation



Hi-C compared to optical map scaffolds

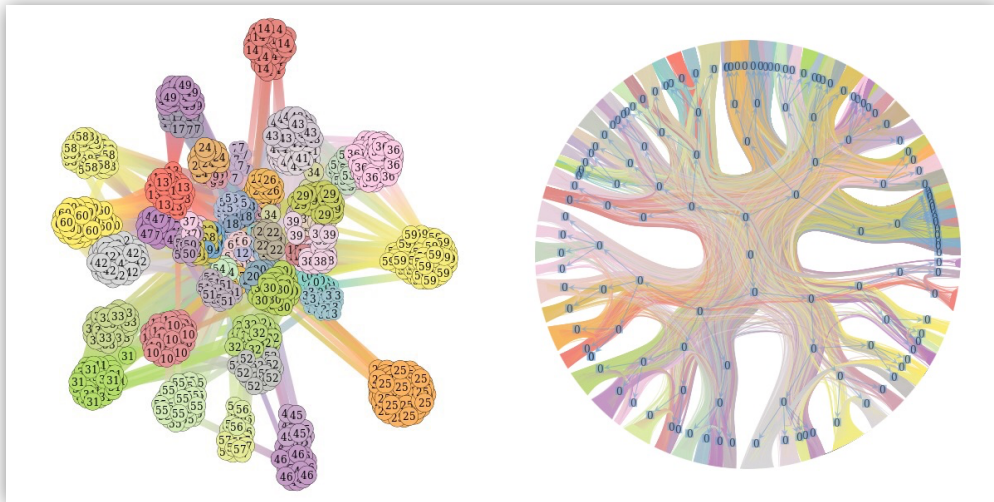
Hi-C based phasing



Hi-C based FALCON-Unzip phasing
<https://github.com/phasegenomics/FALCON-Phase>

Hi-C on important metagenomic samples

- Hi-C can be used to bin PacBio bacterial contigs into genome clusters
- It can also associate plasmids to the host, and enables eukaryotic metagenomics

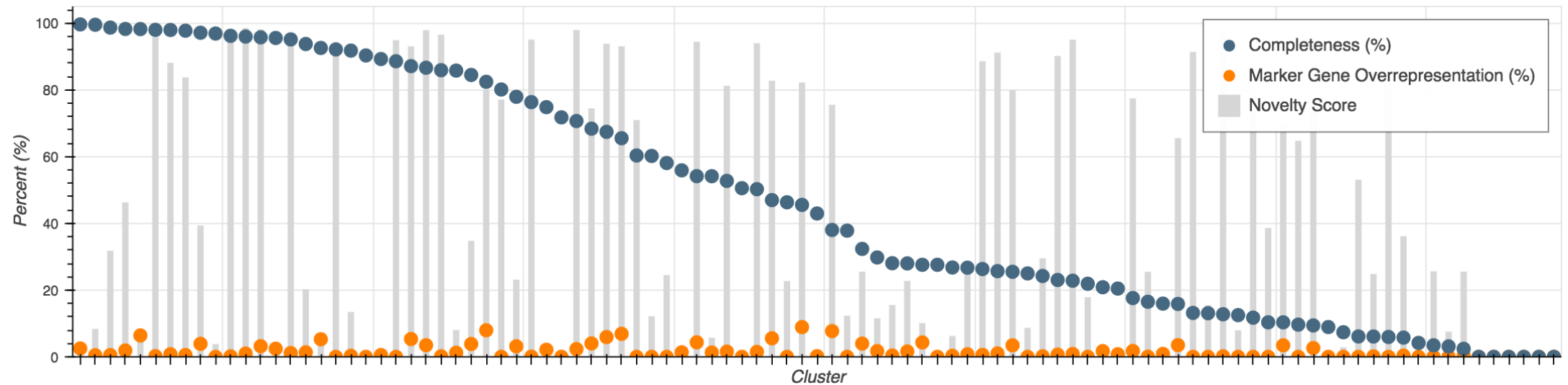


PacBio assemblies + Hi-C = many complete bacterial genomes



ProxiMeta Results

Lil bub
01/13/2018



Conclusions

- Hi-C proximity information is a powerful tool for genome scaffolding
- Hi-C can be used to correct phase switching in PacBio assemblies
- Hi-C has powerful metagenomic applications which complement long SMRT reads

Questions