FALCON-Phase: Phased Diploid Assemblies through Integration of PacBio and Hi-C Data

Zev Kronenberg & Sarah Kingan

Sept 19th
PacBio User Group Meeting
St. Louis
FALCON-Phase is a collaboration between Phase Genomics and Pacific Biosciences
The benefits of FALCON-Unzip

- Contiguous contigs
- Few chimeric contigs
- ~Resolves haplotypes (haplotigs)

The challenges of FALCON-Unzip

- Contigs are not chromosomes
- Haplotig placement
- Phase switching within haplotigs
- Phase switching between haplotigs
Hi-C data augments genome assembly

- Hi-C scaffolding
- Hi-C phasing

Hi-C provides ultra-range haplotypic linkage information
Chromosome-level scaffolds achieved with Phase Genomics Hi-C

Kaniwa

Hayley Hansen, Brigham Young University

Mint

Kelly Vining, Oregon State University
FALCON-Phase corrects FALCON-Unzip phase switching errors between haplotigs

- **Input formats:**
  - FALCON-Unzip
  - High quality Hi-C (quality matters)

- **Output formats:**
  - Psuedo haplotypes (original)
  - Haplotigs + Primary contigs (VGP)

- **This method does not:**
  - Create new sequence
  - Unzip additional regions
FALCON-Phase works on the chromosome scale

- Scaffold phased contigs
- Build a pseudo haplotype pair file (i.e. phased contigs)
- Re-run FALCON Phase
FALCON-Phase is accurate

Example of chromosome level phasing

Benchmarks

<table>
<thead>
<tr>
<th></th>
<th>Zebra Finch</th>
<th>F1 Bull</th>
<th>Puerto Rican</th>
</tr>
</thead>
<tbody>
<tr>
<td>TrioCanu assay</td>
<td>99.4%</td>
<td>96.7%</td>
<td>NA</td>
</tr>
<tr>
<td>SNP assay</td>
<td>91.6%</td>
<td>96.7%</td>
<td>80%</td>
</tr>
</tbody>
</table>
Summary

• Hi-C is critical for genome assembly

• FALCON-Phase solves phasing errors between haplotigs

• FALCON-Phase works at the scaffold scale

• FALCON-Phase is accurate
Availability

• Preprint:
  • https://www.biorxiv.org/content/early/2018/05/21/327064

• Open source:
  • https://github.com/phasegenomics/FALCON-Phase
  • VGP compatibility is on the development branch

• Production code:
  • FALCON-Phase is part of pb-assembly bioconda eco-system
  • https://github.com/PacificBiosciences/pbbioconda
  • https://github.com/PacificBiosciences/pb-assembly
Questions?