iSeGWalker
An easy handling *de novo* genome reconstruction dedicated to small sequences

Benjamin Saintpierre¹, Johann Beghain¹, Eric Legrand¹, Anncharlott Berglar², Deshmukh Gopaul³, Frédéric Ariey¹

1. Genetic and Genomic of Insect Vectors Unit, Pasteur Institute, PARIS
2. Plasticity of bacterial Genome Unit, Pasteur Institute, PARIS

• Benjamin Saintpierre  • 11 dec. 2015
Gaps Filling Problem

Mapping on a reference

(mapping on a reference)

REFERENCE

UNMAPPED READS

no mapping

Gap Filling

- de novo Assembly
- GapFiller, for paired end experiments and small insert (less than 4kb)

Standard computer ?
Single End experiment ?
Structural variations ?
**in silico** Seeded Genome Walker

1. Parameters → READS → SEED
   - Getting the reads with the seed
   - seed seen?
   - New seed
   - Alignment and consens sequence
   - Coverage enough?

2. Getting all consens
   - Final sequence
The Api0 Sequence Project

Plasmodium falciparum Apicoplast Genome

Api0 unknown sequence

~34,270 nt

4 of the ten primer pairs tested

Position: 33,750
The Results

Coverage of the Api0 sequence

Two different Api0 sequences

First mapping exp.

Second mapping exp.

34,272 nt

5′–CCTATTTATAAATTATAGTAGG–3′
5′–CCTACTATATAATTTATAATAGG–3′
Prospects & Acknowledgments

Conclusion & Improvement

- Fast (~1min/kb for a 1 million reads data, on a Intel Xeon with no significant memory consumption)
- Easy to handle, with a very simple « seed-and-extend » algorithm and no installation required
- Repeated region problem, user must watch over the results to avoid chimeric sequences
- Currently using it in the reconstruction from a single end experiment of a virulent Streptococcus genome, where we determine insertions of large sequences (up to 20kb)

Acknowledgments

Genetic and Genomic of Insect Vectors
Kenneth Vernick
Frédéric Ariey
Eric Legrand
Johann Beghain

Plasticity of bacterial Genome
Deshmukh Gopaul
Anncharlott Berglar