Relative positioning of scaffolds: a challenge with new sequencing technologies

Valérie Barbe
CEA/Genoscope
Evry, France
Microbial Genome Sequencing Evolution

Phrap
(www.phrap.org)

Arachne
www.broad.mit.edu

Newbler
www.454.com

2007 2008 2009

Sanger 12X (~40 projects)

454 GS20/GSflx + Sanger 4X
(~20 projects)

454 Titanium + 454 PE GSflx + Solexa
(1 project)

454 GSflx + Sanger 1X + Solexa
(10 projects)

2007 2008 2009
Sanger reads coverage effect on Supercontigs number obtained after assembly

Supercontig number increases as the Sanger reads coverage decreases.
How to improve microbial whole genome reconstruction?

- **In vitro**
  - 454 Paired-end reads (PE) contribution
  - Tail-PCR Optimizations

- **In silico**
  - Ordering Supercontigs with synteny results
  - Ordering Supercontigs with blast results
How to improve microbial whole genome reconstruction?

- **In vitro**
  - 454 Paired-end reads (PE) contribution
  - Tail-PCR Optimizations

- **In silico**
  - Ordering Supercontigs with synteny results
  - Ordering Supercontigs with blast results
Scaffolding comparisons: Sanger VS 454 PE data

454 PE reads significantly contribute to a decrease in Supercontig numbers.
Scaffolding comparisons: Sanger VS 454 PE data

Set up protocols to obtain 454 PE large insert fragments (C. Battaill’s poster)
How to improve microbial whole genome reconstruction?

- **In vitro**
  - 454 Paired-end reads (PE) contribution
  - Tail-PCR Optimizations

- **In silico**
  - Ordering Supercontigs with synteny results
  - Ordering Supercontigs with blast results
Tail-PCR and derived technologies

TAIL-PCR (Thermal Asymmetric InterLaced PCR)

RESDA-PCR (Restriction Enzyme Site-Directed Amplification PCR)
Using Tail-PCR principle for finishing

Supercontig

First PCR run:

Second PCR run:

Sequencing:

- Pool of degenerate primers
- Only one PCR amplification
- Sequencing PCR product directly with the specific primer used for PCR amplification
Improvements of the Tail-PCR technology

<table>
<thead>
<tr>
<th></th>
<th>Project A</th>
<th>Project B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloning gaps</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>Combinatory PCRs</td>
<td>480</td>
<td>144</td>
</tr>
<tr>
<td>PCR Pools</td>
<td>32</td>
<td>18</td>
</tr>
<tr>
<td>Validated PCRs</td>
<td>23 (72%)</td>
<td>13 (72%)</td>
</tr>
<tr>
<td>Failed PCRs</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Remaining PCRs with</td>
<td>27</td>
<td>15</td>
</tr>
<tr>
<td>individual degenerate primers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCRs obtained with</td>
<td>17 (60 %)</td>
<td>13 (87 %)</td>
</tr>
<tr>
<td>individual degenerate primers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloning gaps closed</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>
How to improve microbial whole genome reconstruction?

- **In vitro**
  - 454 Paired-end reads (PE) contribution
  - Tail-PCR Optimizations

- **In silico**
  - Ordering Supercontigs with synteny results
  - Ordering Supercontigs with blast results
How to improve microbial whole genome reconstruction?

- **In vitro**
  - 454 Paired-end reads (PE) contribution
  - Tail-PCR Optimizations

- **In silico**
  - Ordering Supercontigs with synteny results
  - Ordering Supercontigs with blast results
Ordering supercontigs with synteny results

Draft genome Supercontigs

Reference genome

A: Distance determined between two supercontigs in comparison with a reference genome.

Draft genome Supercontigs

Reference genome

B. Dotplot points between the two genomes representing gene correspondences.

How to improve microbial whole genome reconstruction?

**In vitro**
- 454 Paired-end reads (PE) contribution
- Tail-PCR Optimizations

**In silico**
- Ordering Supercontigs with synteny results
- Ordering Supercontigs with blast results
Ordering supercontigs with blast results

Draft genome Supercontigs

Sequence extremities

blast results

Reference genome

Scaffolding with Bambus*

*Pop M. et al. Genome Res. 2004 14: 149-159
Finishing Pipeline

Assembly

Scaffolding using *in silico* analysis
- Synteny analysis with one reference genome
- Sequence similarities

Scaffolding using *in vitro* analysis
- Tail-PCR
- Combinatory PCRs
Perspectives: development and improvement

- **In vitro**
  - 454 PE large insert fragment contribution
  - Tail-PCR: design primer pools in function of genomic GC composition

- **In silico**
  - Ordering Supercontigs with synteny results using multi-genomes
  - Ordering Supercontigs using sequence extremities: blastP results across Uniprot database
Acknowledgments

Finishing Laboratory

Informatics
Claude Scarpelli
Véronique Anthouard
Arnaud Couloux
Carole Dossat
Frédéric Gavory
Simone Duprat

Comparative Genomic Laboratory
Claudine Medigue
Stéphane Cruveiller
David Vallenet
Zoé Rouy
Béatrice Chane-Woon-Ming

Production
Patrick Wincker
Julie Poulain
Adriana Alberti

Sequence Bio-informatics Analysis Laboratory
François Artiguenave
Vanina Castelli
Jean-Marc Aury
Gaëlle Samson

Jean-Louis Petit (Tail PCR optimizations)