High-throughput BAC Fingerprinting For Constructing Physical Maps

Ming-Cheng Luo

Department of Plant Sciences
University of California, Davis
Why still think about fingerprinting when sequencing is easily accessible?

- Vast majority of species will not be sequenced in the near future.

  For those species, contigs of large-insert genomic clones anchored to genetic maps represent a low-cost alternative to genome sequencing that would greatly enhance the accessibility of their genomes for biological research.

- For large genomes, a physical map is prerequisite for genome-wide sequencing.
Development of fingerprinting methods

By population of restriction fragments:

- One 6-cutter (Olson et al. 1986)

- One 6-cutter + one 4-cutter (Coulson et al. 1986; Klein et al. 2000)

- Multiplexing one 6-cutter + one 4-cutter (Ding et al. 1999)

- One type IIS restriction followed by determine of the nucleotide sequence at the cleavage site (Brenner and Livak 1989; Ding et al. 2000)

- Four 6-cutter + one 4-cutter (Luo et al. 2003)
Development of fingerprinting methods

By detection method:

- Staining (Olson et al. 1986)
- Radioisotope labeling (Coulson et al. 1986)
- Fluorescence dye labeling (Ding et al. 1999)
Development of fingerprinting methods

By electrophoresis type

- Agarose gel (Olson et al. 1986)
- Polyacrylamide gels (Coulson et al. 1986)
- Polyacrylamide gel based sequencer (Gregory et al. 1997)
- Capillary sequencer (Luo et al. 2003)
Development of fingerprinting methods

- Multiple digestion
- Multi-color fluorescence labeling
- Capillary electrophoresis
SNaPshot BAC Fingerprinting

Restriction cleavage

5’ G G C C
3’ C C G G

5’ G G C C
3’ C C G G

Xba I
Fluorescent labeling

5’ T
3’ A G A T C

5’ C T A G A
3’

Xba I
Restriction cleavage and fluorescent labeling

**Bam HI**

```
5’  GGGATCCC
3’  CCTAGGG
```

**Eco RI**

```
5’  GAAATTTC
3’  CTTAAG
```

**Xho I**

```
5’  CTCCGAG
3’  GAGCTTC
```
## Characteristics of restriction sites and labeling of fragments

<table>
<thead>
<tr>
<th>Restriction endonuclease</th>
<th>Restriction site</th>
<th>ddNTP</th>
<th>Fluorescent dye</th>
<th>Color of fragment</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>EcoRI</em></td>
<td>G^[AATTC]</td>
<td>A</td>
<td>dR6G</td>
<td>Green</td>
</tr>
<tr>
<td><em>BamHI</em></td>
<td>G^[GATTC]</td>
<td>G</td>
<td>dR110</td>
<td>Blue</td>
</tr>
<tr>
<td><em>XbaI</em></td>
<td>T^[CTAGA]</td>
<td>C</td>
<td>dTAMRA</td>
<td>Yellow</td>
</tr>
<tr>
<td><em>XhoI</em></td>
<td>C^[TCGAG]</td>
<td>T</td>
<td>dROX</td>
<td>Red</td>
</tr>
<tr>
<td><em>HaeIII</em></td>
<td>GG^[CC]</td>
<td>none</td>
<td>none</td>
<td></td>
</tr>
</tbody>
</table>
Other enzyme combination?

Considerations:

- Generate 5’ overhang
- AGCT only
- 6-bp cut?
- Buffer compatibility

Cheap: 7  More expensive: 40  Don’t care: 55

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>G</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcoRI</td>
<td>(212) *</td>
<td>BamHI</td>
<td>XbaI (840)</td>
<td>SalI (1060)</td>
</tr>
<tr>
<td>HindIII</td>
<td>(212)</td>
<td>BglII (1060)</td>
<td></td>
<td>XhoI (504)</td>
</tr>
</tbody>
</table>

* Prices are based on NEB’s list price of 50,000 units at available largest package.
# Selection of enzyme combination

Predicted numbers of restriction fragments in SNaPshot fingerprints of two *Triticum monococcum* and two *T. turgidum* BACs in the range of 50-500 bp

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>116F2 (107.3 kb)</th>
<th>115G1 (128.6 kb)</th>
<th>BAC1 (173.4 kb)</th>
<th>BAC2 (147.6 kb)</th>
<th>Total (556.9 kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>EcoRI</em></td>
<td>31</td>
<td>38</td>
<td>32</td>
<td>32</td>
<td>133</td>
</tr>
<tr>
<td><em>BamHI</em></td>
<td>21</td>
<td>36</td>
<td>53</td>
<td>32</td>
<td>141</td>
</tr>
<tr>
<td><em>XbaI</em></td>
<td>31</td>
<td>47</td>
<td>38</td>
<td>41</td>
<td>157</td>
</tr>
<tr>
<td><em>XhoI</em></td>
<td>26</td>
<td>30</td>
<td>46</td>
<td>23</td>
<td>125</td>
</tr>
<tr>
<td>Total</td>
<td>108</td>
<td>151</td>
<td>168</td>
<td>128</td>
<td>--</td>
</tr>
<tr>
<td><em>HindIII</em></td>
<td>43</td>
<td>51</td>
<td>68</td>
<td>77</td>
<td>239</td>
</tr>
</tbody>
</table>
Fragment sizing w/ ABI3730

• Any5Dye
• Denatured condition (Hi-Di, 95 ºC, 5')
• 36 cm capillary array / 50 cm capillary array
• Liz-500 Size Standard / longer range
Portion of multi-color fingerprinting profile of a BAC clone

- BamHI
- EcoRI
- XbaI
- XhoI
- Liz-500
## Fingerprinting throughput per single sequencer

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Daily</th>
<th>Weekly (7 days)</th>
<th>Monthly (28 days)</th>
<th>Annually (330 days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABI 3100</td>
<td>480</td>
<td>3,360</td>
<td>13,440</td>
<td>158,400</td>
</tr>
<tr>
<td>ABI 3730</td>
<td>2,160</td>
<td>15,120</td>
<td>60,480</td>
<td>712,800</td>
</tr>
<tr>
<td>ABI 3730XL</td>
<td>4,320</td>
<td>30,240</td>
<td>120,960</td>
<td>1,424,000</td>
</tr>
</tbody>
</table>
### Success rate of the SNaPshot based fingerprinting procedure (ABI 3100)

<table>
<thead>
<tr>
<th>Library</th>
<th>No. clones tried</th>
<th>No. clones succeeded</th>
<th>Success rate</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>BamHI</em></td>
<td>58,261</td>
<td>56,187</td>
<td>96.44%</td>
</tr>
<tr>
<td><em>BamHI BiBAC</em></td>
<td>19,730</td>
<td>19,455</td>
<td>98.61%</td>
</tr>
<tr>
<td><em>EcoRI</em></td>
<td>52,914</td>
<td>51,217</td>
<td>96.79%</td>
</tr>
<tr>
<td><em>HindIII</em></td>
<td>58,615</td>
<td>56,803</td>
<td>96.91%</td>
</tr>
<tr>
<td><em>HindIII BiBAC</em></td>
<td>26,125</td>
<td>25,907</td>
<td>99.17%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>215,645</td>
<td>209,569</td>
<td>97.18%</td>
</tr>
</tbody>
</table>
Fingerprints edit and management

- Fragment size-calling
- True fragments vs. background noises
- Batch fingerprint editing
- BAC cross-contamination check and removal
- Fragment frequency analysis
- Dataset management (clones and fragments)
- BAC-marker hybridization data conversion
FP Pipeliner (www.bioinforsoft.com)

GeneMapper (www.appliedbiosystems.com)
GenoProfiler

To edit 100,000 BAC fingerprints:

- input -- GeneMapper exported text file
  ≈ 5 min.

- input -- 3100/3700 samples files (.fsa)
  ≈ 4.5 hr.

Computer: 3.2 GHz CPU, 1.0 GB RAM

http://wheatdb.ucdavis.edu:8080/wheatdb/
A sample contig (≈ 3 Mb)
Fragment sizing

Integrated databases

Contig assembling

Data processing

BAC-marker integration

BACs

DNA isolation

Reactions

http://wheatdb.ucdavis.edu: 8080/wheatdb
Current Applications

Wheat
Rice
Barley
Soybean
Citrus
Sorghum
Brassica
Grape
Tomato
...

Catfish
Rainbow Trout
...

...
Summary

- High information contents
- Accurate fragment sizing
- Simple procedure, high reproducibility
- Relatively inexpensive
- Automated pipeline, high-throughput
Acknowledgements

Jan Dvorak
Carolyn Thomas
Patrick E. McGuire

Olin D. Anderson
Frank M. You

C. Robin Buell
Shu Ouyang
Jo Hsiao

NSF/Plant Genome Research Program

DBI-0077766
“Mapping and sequencing large genome: Let’s get physical!”