Sequences of 14,600 Gene-Bearing Minimal Tiling Path BACs of Morex Barley

- Starting Points
- Combinatorial Pooling Method
- Results
- Information portals
5300 Mb Barley Genome

>80% highly repetitive DNA

~12% "gene space"

Gene-Bearing BACs

Hypomethylated genomic DNA

Shotgun genome sequences

Transcripts


Morex BAC library
(from Clemson University Genomics Institute)

- 313,344 clones
- Average insert size 108 kb
- 6.4 genome coverage

Gene-Bearing BAC Sequencing Strategy

A. BAC library clones arrayed on membranes
B. Gene-bearing BACs identified by hybridization using genic probes
C. HICF fingerprinting and FPC to assemble gene-bearing BACs into contigs
D. Minimal Tiling Path (MTP) computed, MTP clones rearrayed
E. Combinatorial pools of BAC cultures produced, BAC DNA extracted
F. Each pool of BACs sequenced
G. Sequence reads from pools deconvoluted to individual BACs
H. Each BAC assembled using deconvoluted reads
313,344 BACs in 6.3x Morex Library
Yu et al. 2000. TAG 101:1093-1099

hybridizations

83,831 Gene-Positive BACs
Contain ~75% of expressed genes

HICF by MC Luo

FPC Assembly
65,731 BACs in 9394 contigs
~1700 Mb, ~1/3 of genome
http://phymap.ucdavis.edu/barley

14,763 BACs from Minimal Tiling Path
637 combinatorial pools of BACs (7 x 91)
3-decodable shifted transversal design
83,831 Gene-Bearing Morex BACs

UCR Botany & Plant Sciences
Timothy Close
Steve Wanamaker
Kavitha Madishetty
Jan Svensson
Edmundo Rodriguez
Matt Moscou
Josh Resnik
Prasanna Bhat
Pascal Condamine
Harkamal Walia
Hung Le

UCR Computer Sciences
Stefano Lonardi
Tao Jiang
Jie Zheng
Serdar Bozdag
Yonghui Wu

Other BAC Contributors
Andy Kleinhofs, WSU (including Maroof, Gustafson, others)
Nils Stein (Altschmeid, Graner), IPK
Gary Muehlbauer, U Minnesota
Roger Wise, Iowa State U
Muharrem Dilbirli (K Gill), WSU
Lol Cooper (Hayes, Lemaux, Bregitzer)
Catherine Feuillet, INRA
Anders Falk, Uppsala
Tim Sutton & Nick Collins, ACPFG
David Laurie, JI Centre
Leila Feiz (Tom Blake), Bozeman

HICF and Online Physical Map
phymap.ucdavis.edu/barley/
Ming-Cheng Luo, Heather Witt & Yaqin Ma, UC Davis
Frank You, USDA Albany
Gene coverage in a minimal tiling path (MTP)

In an ideal minimal tiling path (MTP), a gene would be covered by one or two BACs

- This ideal case requires two-decodability in BAC pool design.

A computed MTP is very unlikely to be ideal

- We employed a BAC pooling design with 3-decodability

Pooling Design for BAC Sequencing

2197 BACs
91 pools: 7 layers, 13 pools per layer
169 BACs per pool
Each BAC in 7 pools, one per layer
0 BACs shared between pools in a single layer
13 BACs shared between any two pools in different layers
Pooling Design for BAC Sequencing

2197 BACs
91 pools: 7 layers, 13 pools per layer
169 BACs per pool
Each BAC in 7 pools, one per layer
0 BACs shared between pools in a single layer
13 BACs shared between any two pools in different layers

**Layer** | **Pools** | **BACs**
--- | --- | ---
1 | 01-13 | 2197
2 | 14-26 | 2197
3 | 27-39 | 2197
4 | 40-52 | 2197
5 | 53-65 | 2197
6 | 65-78 | 2197
7 | 79-91 | 2197

**Valid Signature**
03,22,34,52,63,67,90

**BAC#**
0296
Pooling Design for BAC Sequencing

2197 BACs
91 pools: 7 layers, 13 pools per layer
169 BACs per pool
Each BAC in 7 pools, one per layer
0 BACs shared between pools in a single layer
13 BACs shared between any two pools in different layers

<table>
<thead>
<tr>
<th>Layer</th>
<th>Pools</th>
<th>BACs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>01-13</td>
<td>2197</td>
</tr>
<tr>
<td>2</td>
<td>14-26</td>
<td>2197</td>
</tr>
<tr>
<td>3</td>
<td>27-39</td>
<td>2197</td>
</tr>
<tr>
<td>4</td>
<td>40-52</td>
<td>2197</td>
</tr>
<tr>
<td>5</td>
<td>53-65</td>
<td>2197</td>
</tr>
<tr>
<td>6</td>
<td>65-78</td>
<td>2197</td>
</tr>
<tr>
<td>7</td>
<td>79-91</td>
<td>2197</td>
</tr>
</tbody>
</table>

Positive Pools
01,03,20,22,30,34,44,52,62,63,67,71,84,90
Pooling Design for BAC Sequencing

2197 BACs
91 pools: 7 layers, 13 pools per layer
169 BACs per pool
Each BAC in 7 pools, one per layer
0 BACs shared between pools in a single layer
13 BACs shared between any two pools in different layers

Layer | Pools | BACs
--- | --- | ---
1 | 01-13 | 2197
2 | 14-26 | 2197
3 | 27-39 | 2197
4 | 40-52 | 2197
5 | 53-65 | 2197
6 | 65-78 | 2197
7 | 79-91 | 2197

Valid Signatures | BAC#
--- | ---
03,22,34,52,63,67,90 | 0296
01,20,30,44,62,71,84 | 1179
Pooling Design for BAC Sequencing

2197 BACs
91 pools: 7 layers, 13 pools per layer
169 BACs per pool
Each BAC in 7 pools, one per layer
0 BACs shared between pools in a single layer
13 BACs shared between any two pools in different layers

Positive Pools
01, 03, 07, 16, 20, 22, 29, 30, 34, 44, 46, 52, 54, 62, 63, 66, 67, 71, 82, 84, 90
Pooling Design for BAC Sequencing

2197 BACs
91 pools: 7 layers, 13 pools per layer
169 BACs per pool
Each BAC in 7 pools, one per layer
0 BACs shared between pools in a single layer
13 BACs shared between any two pools in different layers

Valid Signatures
03,22,34,52,63,67,90  0296
01,20,30,44,62,71,84  1179
07,16,29,46,54,66,82  1861
Results of Combinatorial Sequencing of BACs
- From one set of 2197 BACs, 1 Illumina HiSeq flow cell
- 2167 BACs >20X coverage, mean = 137X
- Assembled using Velvet, mean # nodes = 126

<table>
<thead>
<tr>
<th>DNA Samples Sequenced</th>
<th>Size (Mb)</th>
<th>N50</th>
<th>% Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAC Assembly: # of BACs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (deconvoluted by STD, mean)</td>
<td>0.129</td>
<td>7,210</td>
<td>87.8</td>
</tr>
<tr>
<td>169 pooled (mean)</td>
<td>22</td>
<td>4,270</td>
<td>69.5</td>
</tr>
<tr>
<td>2197 pooled (SOAPdenovo)</td>
<td>286</td>
<td>3,845</td>
<td>56.6</td>
</tr>
<tr>
<td>Whole Genome Assembly (+Ns)*</td>
<td>5300</td>
<td>2,857</td>
<td>30.5</td>
</tr>
</tbody>
</table>

*Whole Genome Assembly is 24X by SOAPdenovo, includes ~90% of all known barley genes (BLASTN e⁻²⁰); Barley Genetics Newsletter 41, poster PO114.
Outcome is sets of unordered sequences allocated to bins defined by MTP BAC ends
BLAST Morex MTP BAC Sequence Database
www.harvest-blast.org
Retrieve Morex MTP BAC Sequences
www.harvest-web.org/utilmenu.wc
Anchoring the Physical Map to the Genetic Map

BAC contigs in physical map represented by minimal tiling path (MTP) + SNP Markers on linkage groups of genetic map = BAC contigs positioned on genetic map
Relating Genes to BACs Using Illumina GoldenGate Assay

Gene present: Genotype Call

- A/A
- A/T
- T/T

Gene absent

null
Positive controls, Morex genomic DNA

Negative controls, *E. coli* DNA

BAC pools positive for this gene

Positive controls, Morex genomic DNA

• 7 pools are positive for the SNP gene (blue dots), so this gene is contained in one BAC among 2197 MTP BACs contained in 91 pools
HarvEST: Barley - Windows version (from harvest.ucr.edu)

- Display below shows barley-rice synteny
- Includes BACs anchored to SNP loci (shaded green)
- Adding new BAC and sequence export functions now
- Anticipating deep links between HarvEST and MIPS/IPK database, data to GrainGenes for TriticeaeCAP portal, etc.
Sequences of 14,600 Gene-Bearing Minimal Tiling Path BACs of Morex Barley

UC Riverside, Department of Computer Sciences
   Stefano Lonardi, Denisa Duma, Matthew Alpert, Yonghui Wu, Gianfranco Ciardo, Burair Alsaihati

UC Riverside, Department of Botany & Plant Sciences
   Prasanna R. Bhat, Yaqin Ma, Steve Wanamaker, Josh Resnik, TJ Close

University of Torino, Italy
   Francesca Cordero, Marco Beccuti

UC Davis
   MingCheng Luo

Thank You For Your Attention