



United States Department of Agriculture  
National Institute of Food and Agriculture

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

- Starting Points
- Combinatorial Pooling Method
- Results
- Information portals

Timothy J Close  
UC Riverside



United States  
Department of  
Agriculture

National Institute  
of Food and  
Agriculture

# 5300 Mb Barley Genome

>80% highly  
repetitive DNA

~12%  
“gene  
space”

Gene-  
Bearing  
BACs

Hypomethylated  
genomic DNA

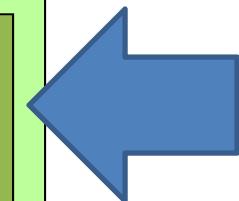
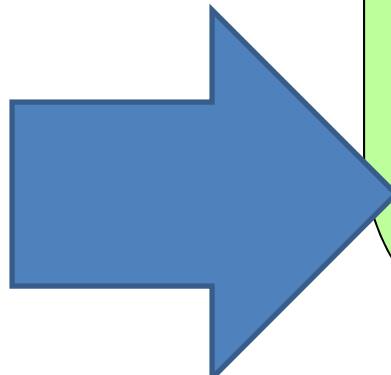
Transcripts



Wicker et al.  
2005. Plant  
Journal 41: 184-  
194

Barakat et al.  
1997. PNAS 94:  
6857-5891

Shotgun  
genome  
sequences

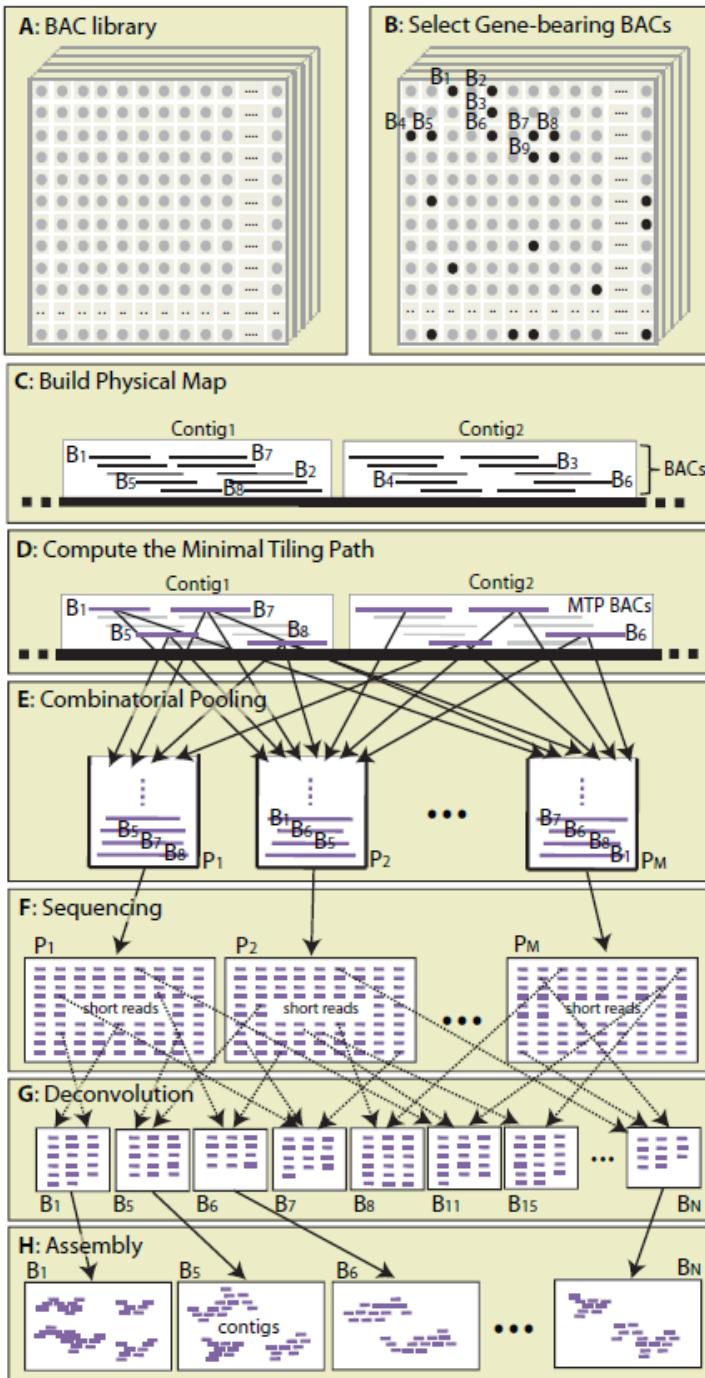


## Morex BAC library (from Clemson University Genomics Institute)

---

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

Yu Y, Tomkins JP, Waugh R, Frisch DA, Kudrna D, Kleinhofs A, Brueggeman RS, Muehlbauer GJ, Wise RP, Wing RA. 2000. A bacterial artificial chromosome library for barley (*Hordeum vulgare* L.) and the identification of clones containing putative resistance genes. TAG 101:1093-1099



## 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 Dilbirligi (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

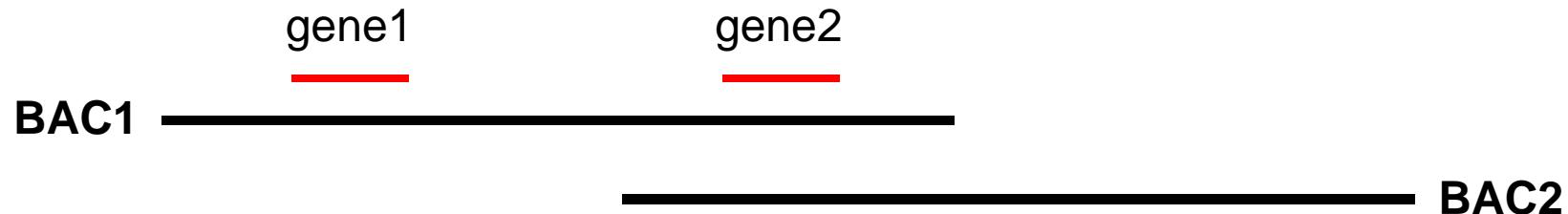
[phymap.ucdavis.edu/barley/](http://phymap.ucdavis.edu/barley/)

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

Thierry-Mieg, N. 2006. A new pooling strategy for high-throughput screening: the shifted transversal design, *BMC Bioinformatics* 7:28

# 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

# Pooling Design for BAC Sequencing

2197 BACs

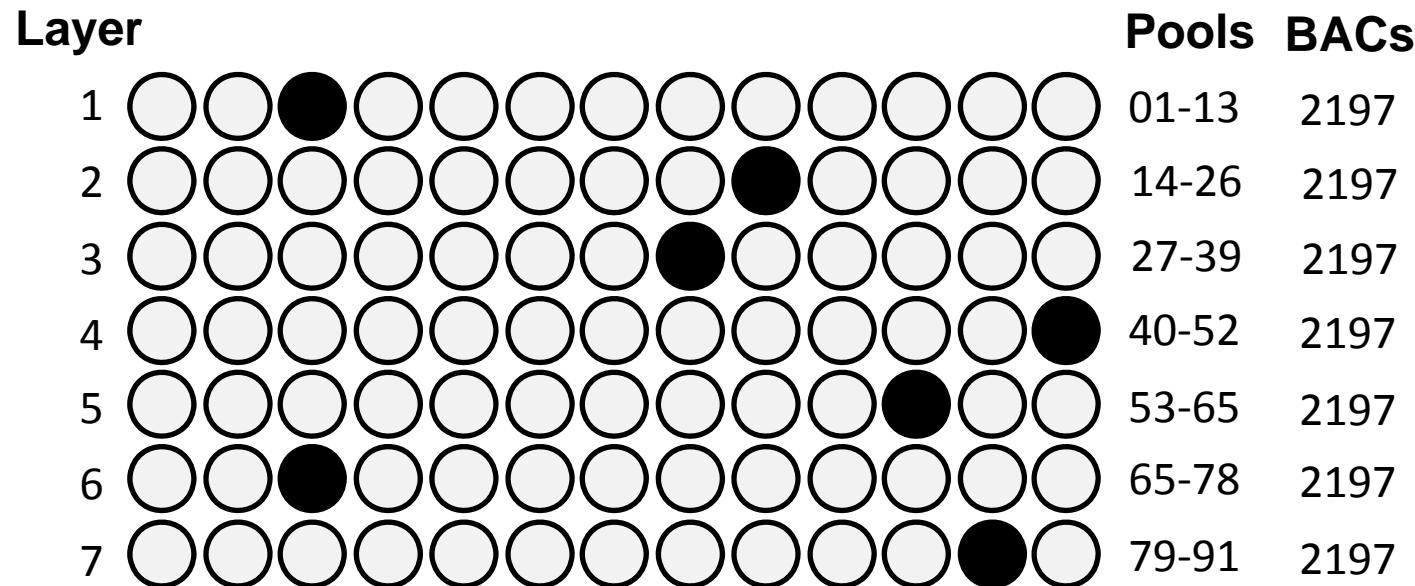
91 pools: 7 layers, 13 pools per layer

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 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

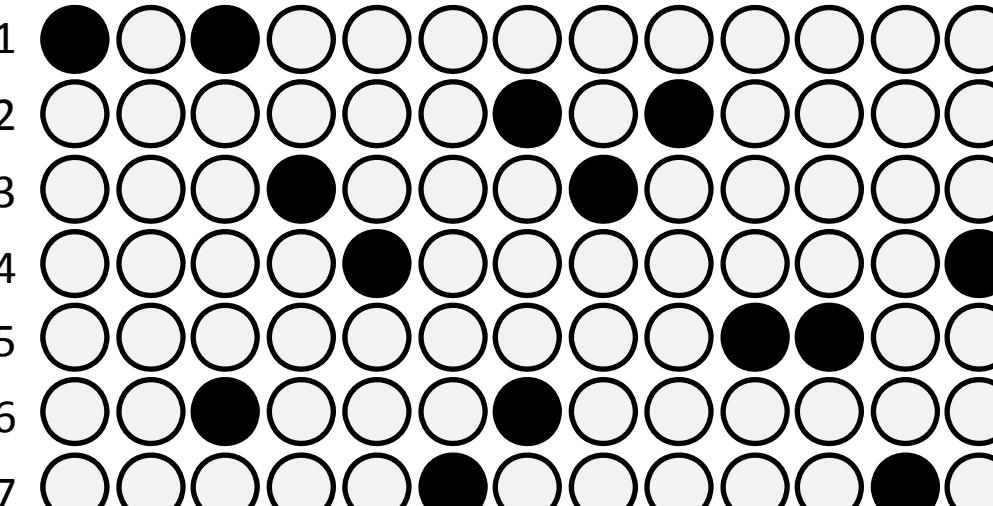
169 BACs per pool

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

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



## Positive Pools

01,03,20,22,30,34,44,52,62,63,67,71,84,90

# Pooling Design for BAC Sequencing

2197 BACs

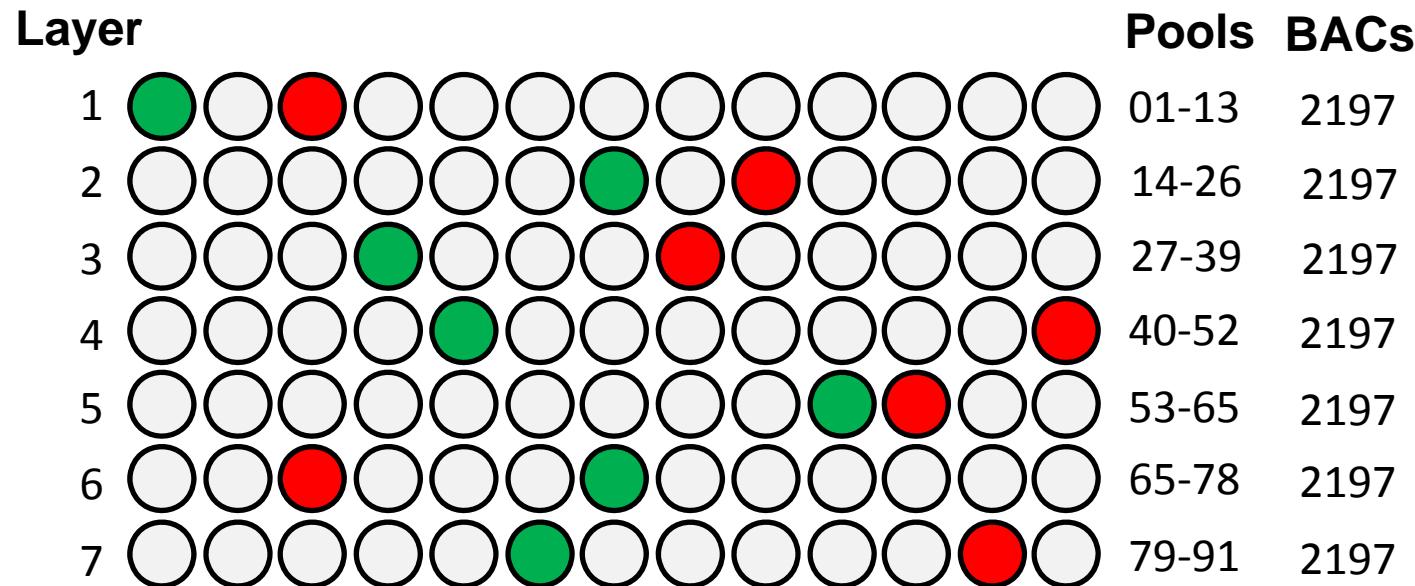
91 pools: 7 layers, 13 pools per layer

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



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

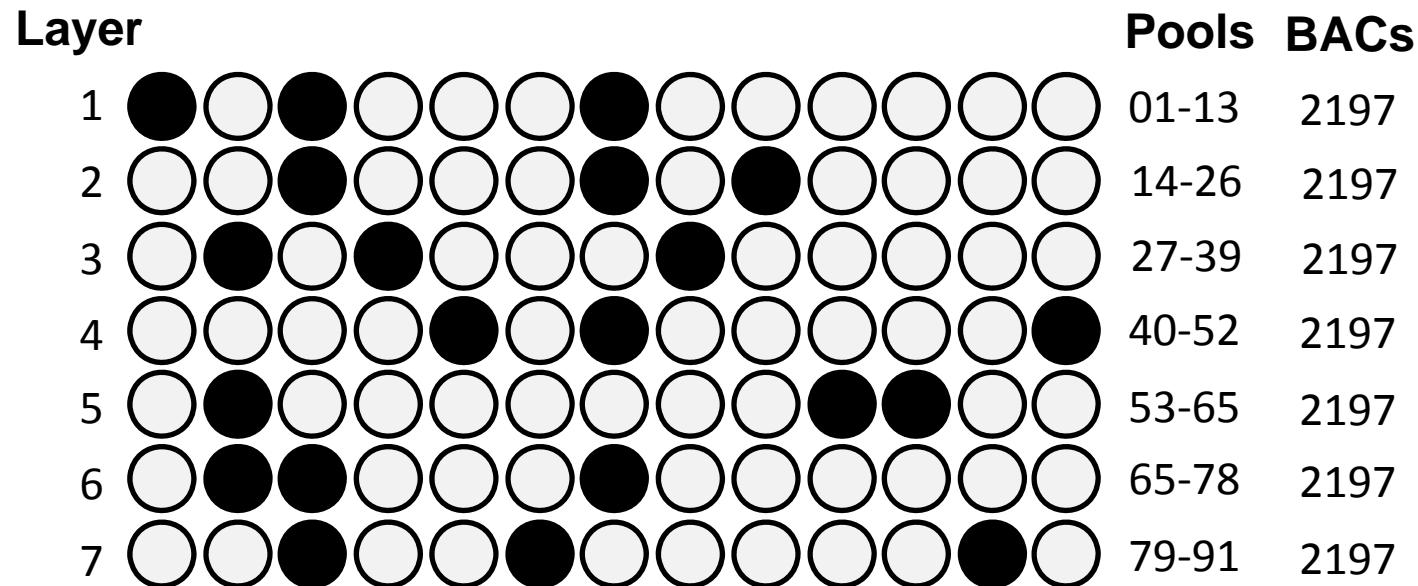
91 pools: 7 layers, 13 pools per layer

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

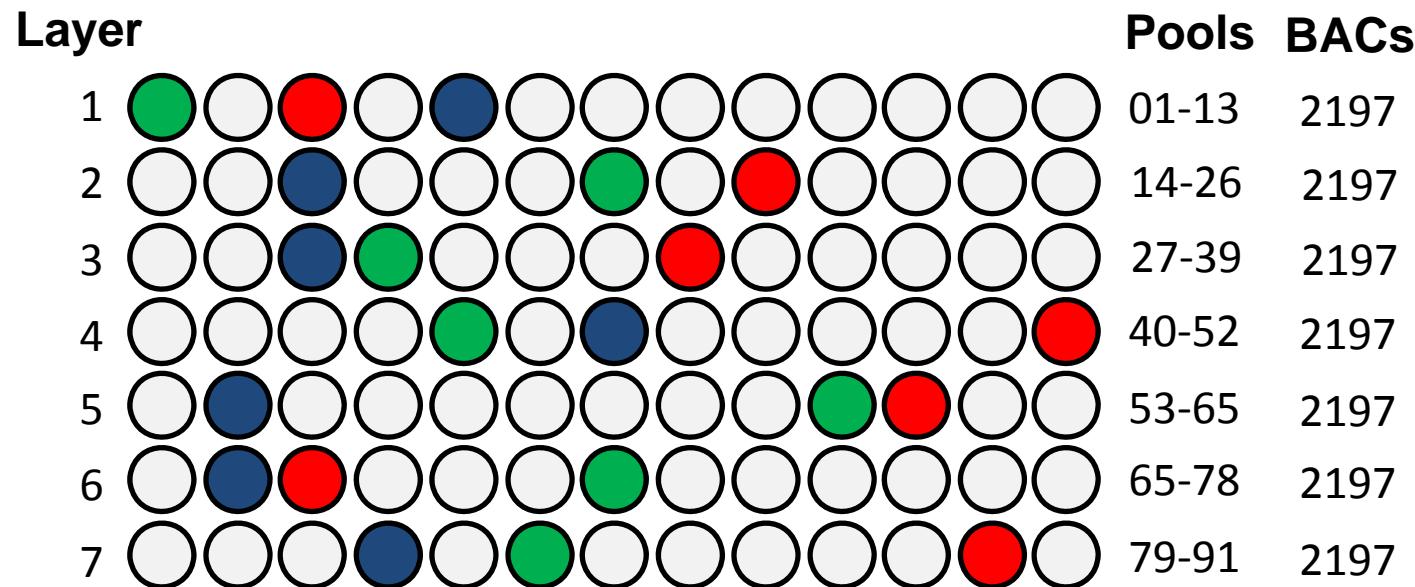
91 pools: 7 layers, 13 pools per layer

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	BAC#
03,22,34,52,63,67,90	0296
01,20,30,44,62,71,84	1179
07,16,29,46,54,66,82	1861

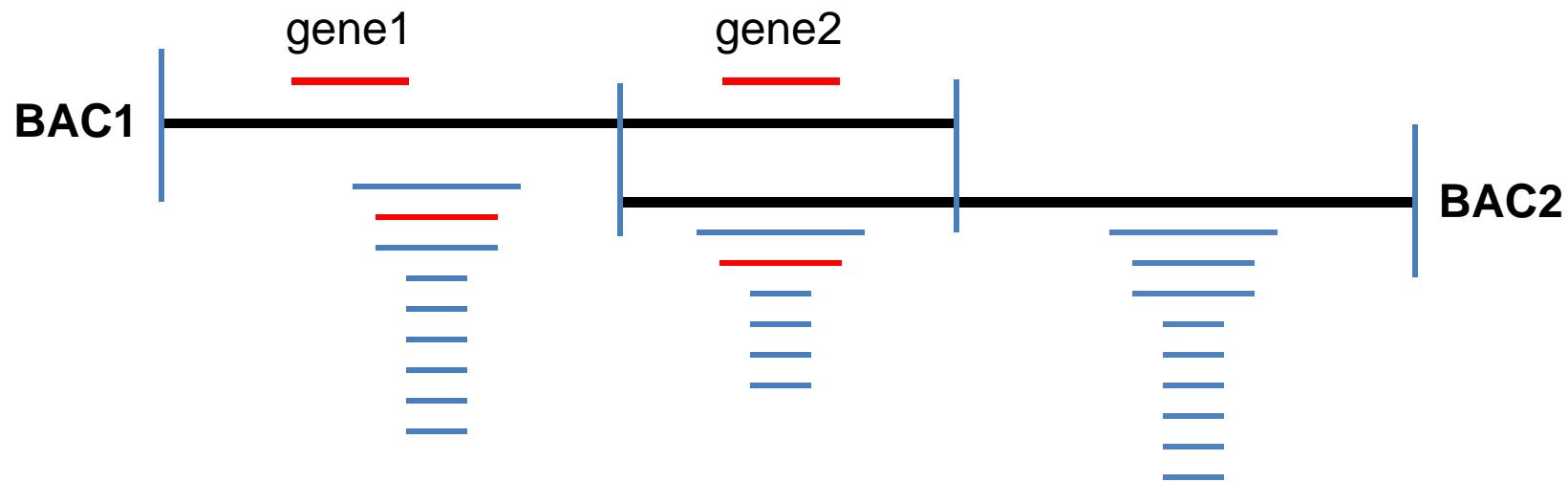
## Results of Combinatorial Sequencing of BACs

- From one set of 2197 BACs, 1 Illumina HiSeq flow cell
- 2167 BACs >20X coverage, mean = 137X
- Assembled using Velvet, mean # nodes = 126

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

\*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

The screenshot shows a Mozilla Firefox browser window titled "HARVEST-BLAST.ORG - Mozilla Firefox". The address bar displays "HARVEST-BLAST.ORG". The main content area is titled "HarvEST BLAST Search". A sub-section titled "Choose program to use and database to search:" is present. Under "Program", "blastn" is selected from a dropdown menu. Under "Database", "Morex Barley BACs" is selected from a dropdown menu, which is also a list of available databases. The list includes:

- Morex Barley BACs
- Barley 2943 Mapped SNPs
- Barley 21 (Affymetrix Chip)
- Barley 25
- Barley 31 stringent
- Barley 32 relaxed
- Barley 35 relaxed
- Barley Genome 0.05
- Morex Barley BACs
- Barley\_Transcriptome\_0.02 (h21,Conf 0.5) 59200
- Barley\_Transcriptome\_0.02 (h21,low conf) 11571
- Brachypodium B01 stringent
- Brachypodium B02 relaxed
- Citrus C37 (Affymetrix Chip)
- Citrus C38 relaxed
- Citrus C52 relaxed
- Citrus C53 relaxed
- Citrus C54 relaxed
- Citrus C55 relaxed
- Coffea J01 stringent

Below the database selection, there is a sequence input area containing a FASTA sequence:

```
ACGTGTGTGACGATCGTACCGAGCT  
ACGAGCTGTCAAATCGATCGACCT  
TCGACCTTTGACGTGTGACGCA  
GTGACGATCGTACGAGCTGTCAAA  
TGTCAAATCGATCGACCTTTCGAC  
TTTCGACGTGTGACGATCGTAC
```

Below the sequence input, there are fields for "Or load it from disk" (with a file input field), "Set subsequence: From" (with a text input field), and buttons for "Clear sequence" and "Search".

# Retrieve Morex MTP BAC Sequences

[www.harvest-web.org/utilmenu.wc](http://www.harvest-web.org/utilmenu.wc)

**HarVEST: Utilities Menu - Mozilla Firefox**

File Edit View History Bookmarks Tools Help

HarVEST: Utilities Menu

harvest-web.org/utilmenu.wc

Google

## HarVEST: Utilities

### Get a FASTA File of Selected Morex Barley BACs

Please paste your list of identifiers or upload a query file that contains identifiers.

The file can contain one-line descriptions from BLAST output like this:

```
203B04_NODE_0001
203B04_NODE_0004
```

Your list can contain only BAC addresses (like 203B04) if you select "Download the whole BAC(s)" below.

Paste Your List Here

```
203B04
```

Or Upload a Query File  Browse...

Download the whole BAC(s)

End each line with CR+LF (DOS/Windows)  
 End each line with LF (Unix/Linux)

Submit Query

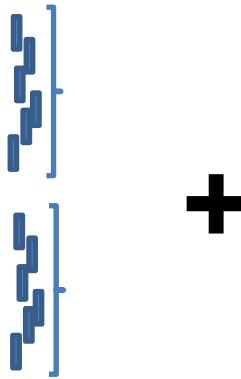


**MOREX\_HV5.txt - Notepad**

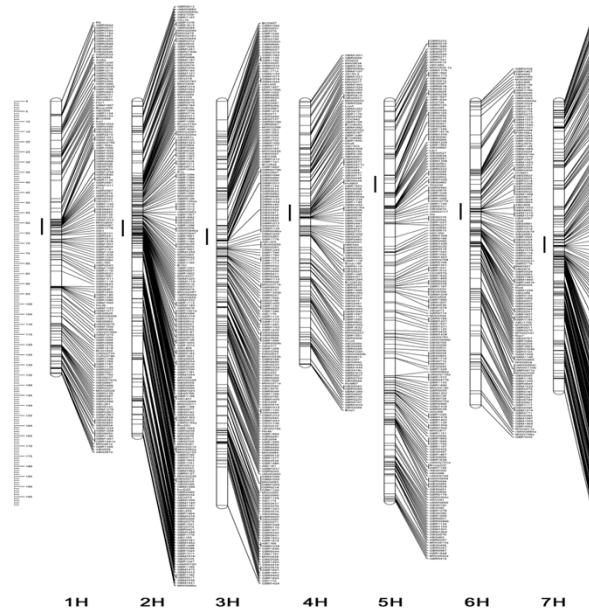
File Edit Format View Help

```
>203B04_NODE_0001
ATCTCTTCAATTGTTCCGCTATTAGTGACTTGGAGTGACTCTTGTTCAGTGGTAGGCTAGTTATGAAAGTAT
GAACCTTAGGCCCTGTTCACGCGTCAATACGGTCGTTCTCGCTTTTATTACT
TACCTTGCTGTTTGATTTGAGTTTACAGTTACAAAACCTATATCTACCATCCATATTGCAC
TTGTATCACCACATCTTACCGAAGCTAGTGACCTATACTATTTACCATTTGATTGGGTG
TGTTGGGACACAAGGACTCTTGTATTGGTTCAGGGTTGCTTGAGACAGACCATC
TTCATGCTACGCCCTCCGGATTGATAACACCTAGGTCAACCCATTGAGGGAAATTGC
TACTGTCTCACAAACCTCTGCACATGGAGGCCAACACAGTCTACAAGGAGAAGGTTGCA
TAGTAGGCATACGCACCGAGAGCATCCAGGTACTCTTCGGGCCCTTGTTCAGC
CGCTCCCTGAGGCCCTGCTGGACCTACCTCTCCCTCTCCATGGCATTCCGAG
CCCATAAATCTCTCTCGGACCTGCAACGGCCGATGCCGAGGGTACCTTTTG
GCCCGCTCCAATCGGGGTCTCCGATCCCCCTTTCCCTTGAAGAACGACCAAGGTTG
TCTCCTCATGCCGACCGCCGTGCTTCTCAACCATTGGTACCCGGGACT
ATCCACCGCGACTGGTCAACCTGCGACTCTGGACACTGAGGGGGCTCTCTTAGGAG
CTTCCGGTGGAGCTAGGCCAGGGAGTGGCTCTCCATGGCATTCCCTGAGCATGCTGGT
GCCGCGCGTCAACAGCCATACCGGAACGACAGGGCAGGGCACGGCTAACCTGGTCC
CTCTCTCCCAGCTCGTCCGATGGGGAGGTAGGGACGACGTGATGGCTGGGCC
CCGCTGTAAGCGCCTAGGGCGCTCCCTCGGTTAAGTGGAGGCTGGTCTCCCTGA
AGCCGCTCCCTGCAAGGATTGGTTCAGCTAGTGGCTTCAAGCTGGCCCTCGC
TGGCAGGACGAGCACTTGGCGGCCACCTCTCTCCAGCGCCGGGCACAGTTAGC
CCCCCTAGATAACTCCCTTCTCCCTTCTCTCATAGAGATGCAAAATACTCATAACT
ACTAAAGTAACTCCAAATCATTGATTTAAATCTACAGGTCCTAAATCCTTGCC
TATCCATATCTGTGCAATTGACATTTTACAAGAATTTTCACTAGTATTATTCAA
ATGCAAATTAGTATTGAAACTCCACCATGTAACCTCTTGAGCCTCAAATCAACT
CCAAATGAGTCTCAATTCTAGGTTATAAAATCATTCCCTAAATTCTTAGCCT
CGGTGAAACATTCTGCTATATTCTGGTACATCAAATAATAGCCCTTTTCTAT
TTGTTCCATATTAGAAAATCATAGGAAATTCAACCAACTCCAAAACAGTGAACAC
TTCGAATGTTCTAAATTCTGCTATTAAATGAGTGGCTCACTCATTTTAAAG
AAAGTTCTGTAAGCTTATAGGACCTTACTCTTCTAAATTCAAATTCAAAATC
TCAATGATCCAAATCTCTAGAAGAACGTTAGTATTCTTATGACCAGAGGAGC
AAGAAAAAAATTGACTTTAGGCGCTTGTCTGCTGTTGAGCTTACGGTGG
TTGTTCCGACGATAGTGGCAAATAGCGGAGTACTGGAGTTGAGTAAAGACCTT
GAAGACCTAGGAGACCTACTGGCAAGGCAAGCCCTTGACCATGCTGAGGAAT
ATTGGTGCATGCCATTACTTACTCCGATGCAATGACTATCGTAAACCGT
CGATATGGTGTACCTATGGCTCTGGAGCACTTGAATTGAGCATGACCTTACATA
TTGAGGGTACCTAGGAAAGTGAACAGTAGATAGTGGTACCTGGATAG
CATATGGATGGTGTGACTCAAAAGGTGATATCAAAGGTTTCAAAACATTGCGG
GTGCACTAAAGCTGAGAGTGTGTTGAGTGGCTGACCACTTGAGAAAGAATTGG
TTTACCGCAAAAGGTTGTGAGTGGTTAGGAATGGGATTAGTTAAAGATTGCTG
ACCATCCCAACCTTACATGCGACCCAGTACCTTTATGGGAGGGGTTTGTGAT
TACTGTGTCGACTAGGCAAGAGGCCATTACAAAGTGGGGAGTAATATGGTACTC
TCGTGACGGGGTGTGCAAGTTAGGGTACTATGCTGTTTGACCTGTTCCGGCACACC
ATTGGTCCCGCATGATAATACGGTCTAGAGTGGTGTGCTGAAGATGACATCGTG
GGCTGGTACAAATAGAGTGGACTTTTCTAGTGTGTCACGAGAAAAGGCAATGATCGG
GGTACTACCTCGGGTATGGTATATACCGCGAGTCAGGATGAGAAGGAAGTTCGGGA
ACTCGTGGGTACCGTATGACCTCTGCAGCGTACGATCAATGACAGGTGGTAATCCCG
CTTAGAGTATTCCACATAACCAAGTGTGTTGTTGATAAAAGGTGTTACATAAGAAGT
GAATACCTGGTCAAGTATGGCAGCTGACATGGAGGGTGAACATGAGGTGTTCAACCC
CAGAATATTGTTGATATGTAACCTTACGGTTACTATCTACCTTATGATGC
```

# Anchoring the Physical Map to the Genetic Map

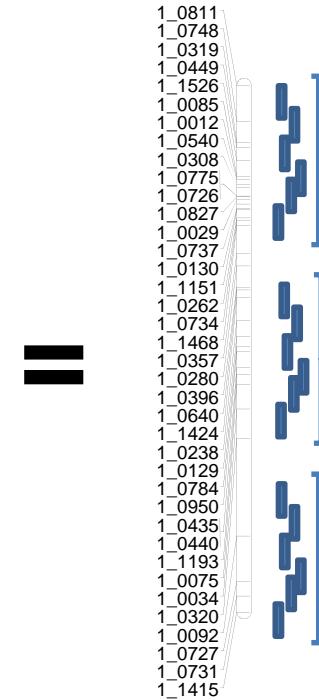


+



BAC contigs in  
physical map  
represented by  
minimal tiling  
path (MTP)

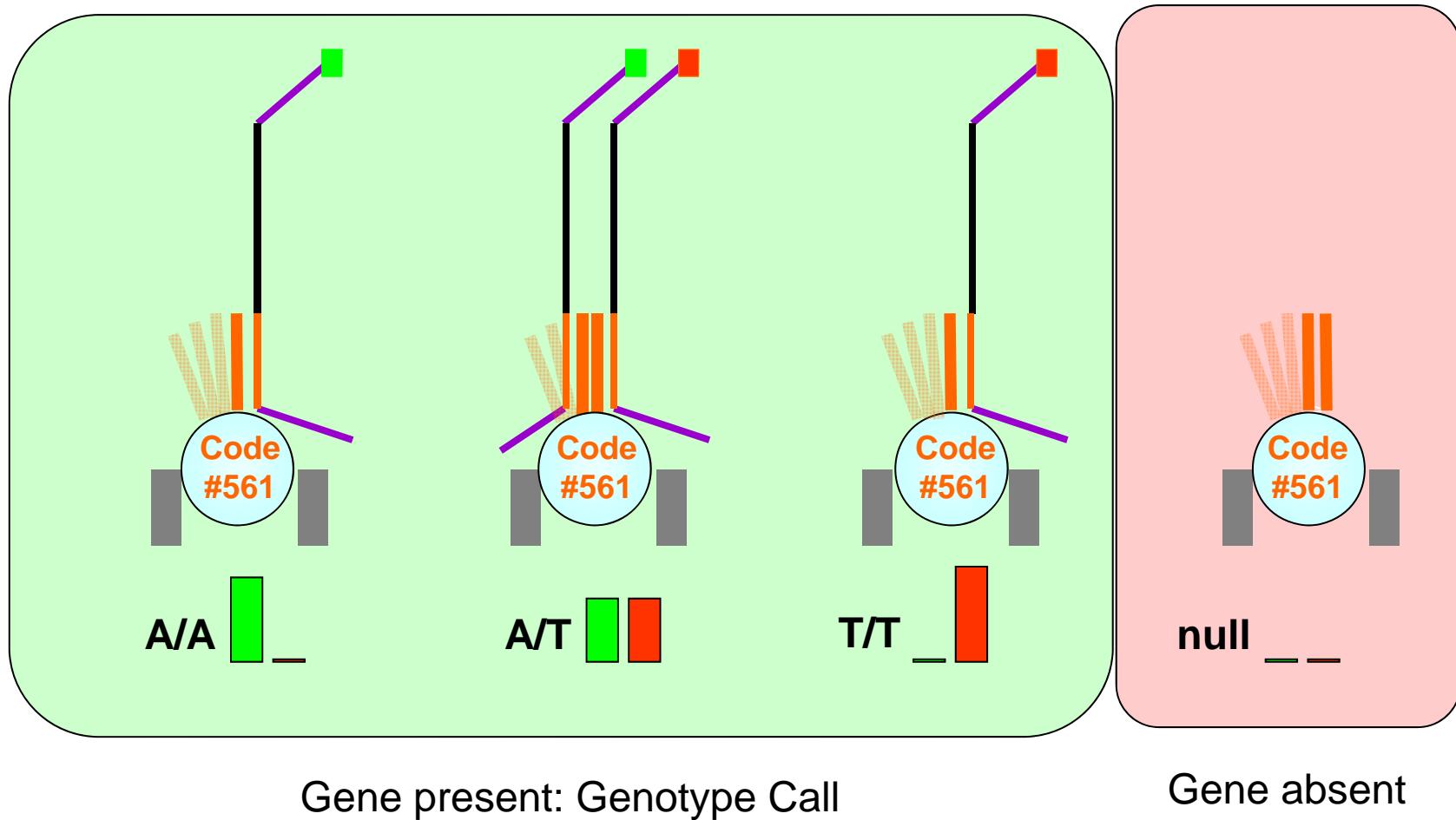
SNP Markers on  
linkage groups of  
genetic map



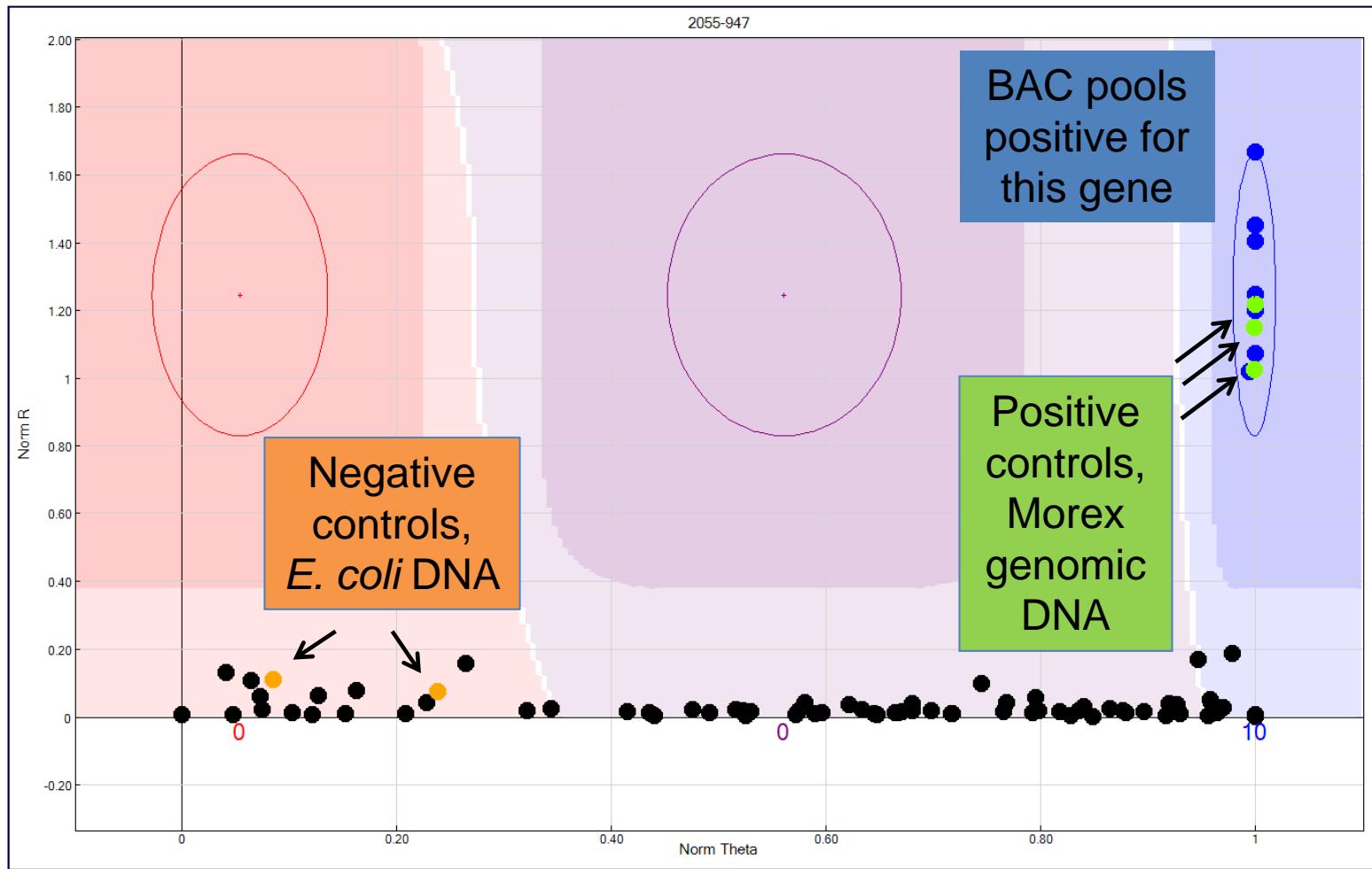
BAC contigs  
positioned on  
genetic map

# Relating Genes to BACs Using Illumina GoldenGate Assay

---



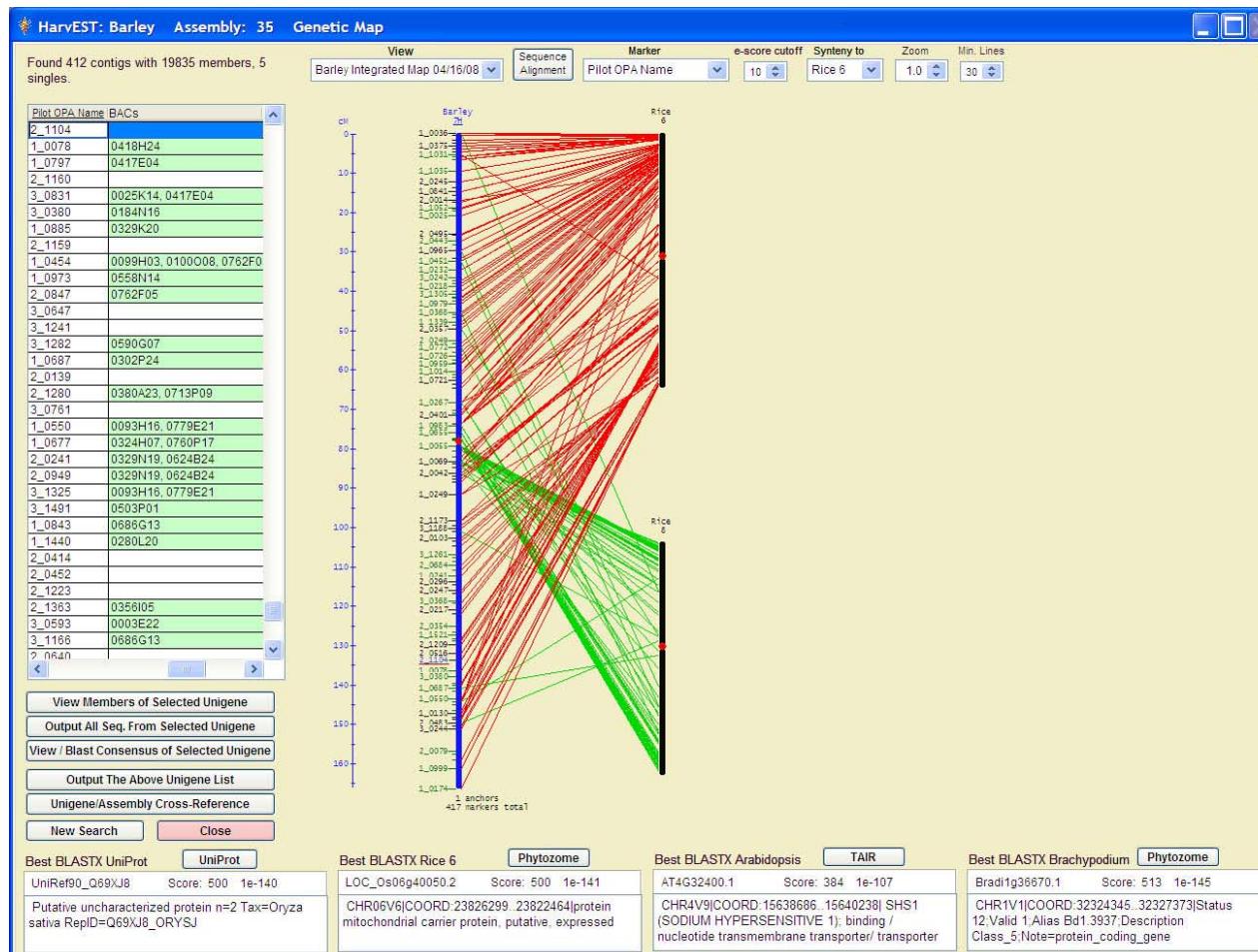
# GoldenGate Assay for Gene Detection in BACs



- 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**