Genomic tools enhance power and precision of hazelnut breeding

Vidyasagar R Sathuvalli, Shawn A Mehlenbacaher, Brooke C Peterschmidt

Department of Horticulture, Oregon State University, Corvallis OR 97331
Overview

- Hazelnut breeding

- Eastern filbert blight (EFB) disease
  - Disease resistance breeding – Marker assisted selection
  - Sources of resistance to EFB
  - Map-based cloning of EFB resistance ‘Gasaway’ gene

- Sporophytic incompatibility

- Development of SSR markers from hazelnut transcriptome
Hazelnut Breeding
European Hazelnut
(Corylus avellana L.)

- Origin - Europe and Asia minor
- Family: Betulaceae
- National Clonal Germplasm Repository
  - 825 accessions of Corylus
  - 429 of C. avellana
- ~ 100 more accessions at OSU

[Images of hazelnut trees and nuts]
Hazelnut Production (MT)

1. Turkey 500,000 65.2%
2. Italy 104,900 13.7%
3. United States 42,640 5.6%
4. Azerbaijan 30,430 4.0%
5. Georgia 21,800 2.8%

Oregon’s Willamette Valley – home to 99% of the U.S. hazelnut industry
Hazelnut Breeding Objectives

A. Blanched kernel market (for chocolate, baked goods)

1. Bud mite resistance
2. Round nut shape
3. High percent kernel
4. Precocity
5. High yield
6. Easy pellicle removal
7. Few defects
8. Early maturity
9. Free-falling nuts

B. Resistance to eastern filbert blight (EFB)

1. Simply inherited resistance (‘Gasaway’ & others)
2. Quantitative resistance (e.g. ‘Tonda di Giffoni’)
Hazelnut Breeding Flow Chart

1. Choose parents, make crosses
2. Grow in greenhouse
3. Seedlings in field
4. "
5. "
6. Evaluate a few nuts
7. Evaluate nuts
8. Layer, Evaluate nuts
9. Nursery, Evaluate nuts
10. Plant replicated trials
11. Trees in field
12. Evaluate nuts
13. "
14. "
15. "
16. Evaluate nuts, Summarize data
17. Release new cultivar

Breeding cycle: 8 years from seed to seed

Mehlenbacher, 2011
Obtaining Hybrid Seed
Growing Hybrid Seedlings (4000 planted per year)
Evaluation of Original Seedlings
Stage-I with removal of discards

Mehlenbacher, 2011
Harvested layers are weak. They are held in the nursery for one year, and then planted in the orchard.
Second Stage of Evaluation
in Replicated Trials
Eastern Filbert Blight in Hazelnut

- Caused by ascomycete *Anisogramma anomala*
- Native to eastern North America
- First discovered in Willamette Valley – 1986
- 16 – 24 months between inoculation and canker development
- More than 60% of Oregon’s hazelnut orchards are affected or in close proximity to diseased orchards
Host genetic resistance was first identified in ‘Gasaway’

An obsolete pollinator

Resistance controlled by dominant allele at single locus

RAPD markers closely linked to resistance were identified

Markers assisted selection (MAS) is being carried out for ‘Gasaway’ resistance
Marker Assisted Selection (MAS)

- Extract DNA from ~3000 seedlings per year, 192 per day
- Amplify using flanking RAPD markers 152-800 and 268-580
- Seedlings that lack markers are discarded
New sources of resistance to EFB

EFB Resistance Sources

Corylus avellana

Quantitative Resistance

Gasaway* Zimmerman 408.040* 759.010* Ratoli* Culpla 495.072 Moscow Selections (N01,N02,N26,N27,N37) Crvenje Uebov COR 187 Amarillo Tardio

Qualitative Resistance

USA Georgia Spain Russia Serbia Finland Chile

*Markers linked to resistance were identified
Assignment to Linkage Groups
(based on co-segregation with SSR markers)
### Assignment to Linkage Groups

<table>
<thead>
<tr>
<th>Variety</th>
<th>LG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gasaway</td>
<td>LG6</td>
</tr>
<tr>
<td>Ratoli</td>
<td>LG7</td>
</tr>
<tr>
<td>Georgian 759.010</td>
<td>LG2</td>
</tr>
<tr>
<td>408.040</td>
<td>LG6</td>
</tr>
<tr>
<td>Culplas</td>
<td>LG6?</td>
</tr>
<tr>
<td>Russian 495.072</td>
<td>LG6?</td>
</tr>
</tbody>
</table>

Are the R genes from ‘Gasaway’ 408.040, ‘Culplas’ and 495.072 the same or part of a cluster?
Map-based Cloning of the EFB Resistance Gene from ‘Gasaway’

- Requirements of map based cloning
  - A large mapping population
  - Genetic map
  - Recombinants between flanking markers
  - A Bacterial Artificial Chromosome (BAC) library
  - Probes to screen BAC library for chromosome walking
In 2007, controlled crosses between OSU 252.146 and OSU 414.062 generated 1488 seedlings

- 07001 – 1080 seedlings
- 07002 – 408 seedlings

07002 is from reciprocal cross, OSU 414.062 x OSU 252.146
Fine Scale Genetic Mapping

- DNA was extracted from 1488 seedlings in 2008
- Screened for the presence of robust RAPD markers 152-800 and 268-580
- Both markers present – Assumed to be resistant
- Both markers absent – Assumed to be susceptible
- 87 recombinants (one marker present and the other absent) were identified – inoculated in the greenhouse
- Resistance is flanked by two RAPD markers W07-375 and X01-825
BAC Library of Hazelnut

- Constructed for ‘Jefferson’
- Average insert size – 117kb
- Genome coverage – 12x
- 39,936 clones arrayed in 104 384-well plates

Sathuvalli and Mehlenbacher (2011) Genome 54:862-867
Screening of BAC Library

Pooling and Screening by PCR

Plate pools - 104
Row pools - 16 per plate
Column pools - 24 per plate

Extract DNA

Screen Plate pools

Screen Row and Column pools

Identify the positive clones

Sequence the BAC ends

18 SCARs from 9 RAPD markers

Screening twice helps avoid false positives and other PCR artifacts

Hit:78-Ex17
Sequence the BAC ends

Design new primers from BAC ends

Map the new markers

Screen the BAC library with new BAC end probes

Identify new BACs

Identify the BACs carrying resistance

3 rounds of screening identified 93 BACs
Markers from BAC end Sequences

- Few polymorphic BAC end markers – a major constraint
- Markers were developed from BAC end sequences
- 41 markers were mapped to the resistance region
  - 23 - Sequence Characterized Amplified Region (SCAR)
  - 7 - Single Stranded Conformational Polymorphism (SSCP)
  - 4 - Simple Sequence Repeat (SSR)
  - 7 - High Resolution Melting (HRM)
Fine Mapping

- High density map
  - 41 BAC end markers
  - 8 RAPDs
  - 2 HRMs from RAPDs
  - Resistance phenotype

- Map spans 4.45 cM

- Averages 0.03cM between markers

- 34 markers placed < 1cM from resistance

Sathuvalli and Mehlenbacher, in preparation
BAC Fingerprinting

- Each BAC screening provided more than one BAC hit
- HICF to merge BACs into contigs

FPC output showing the BACs in a single contig
Physical map of eastern filbert blight resistance region

Sathuvalli and Mehlenbacher, in preparation
Identification of Resistant Contig

- A single recombination event was observed between W07-375 and resistance.
- A single recombination event was observed between resistance and HICF13.
- W07-375 and HICF13 are from the same contig.
- Contig size is ~150kb.
- Resistant contig consists of 3 overlapping clones.

Resistance region

+ Present
- Absent

Sathuvalli and Mehlenbacher, in preparation
Sequencing of Whole BACs

- BACs in the resistance region were sequenced using Illumina technology
- Velvet and SOPRA programs were used to assemble Illumina reads into contigs
- CodonCode Aligner was used to align, trim and correct the contigs from Velvet and SOPRA

Sathuvalli and Mehlenbacher (2011), in dept. review
## BAC Sequencing

### Sequence coverage: 60-100%

<table>
<thead>
<tr>
<th>BAC clone</th>
<th>RAPD marker</th>
<th>Illumina system</th>
<th>Barcode</th>
<th>No. of contigs</th>
<th>Longest contig (bp)</th>
<th>Shortest contig (bp)</th>
<th>Total length (bp)</th>
<th>Estimated size (bp)</th>
<th>Coverage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>49A4</td>
<td>H19-650</td>
<td>HiSeq</td>
<td>GAC</td>
<td>12</td>
<td>34,129</td>
<td>1,833</td>
<td>151,300</td>
<td>125,000</td>
<td>100*</td>
</tr>
<tr>
<td>56N1</td>
<td>152-800</td>
<td>HiSeq</td>
<td>CGC</td>
<td>6</td>
<td>53,963</td>
<td>3,150</td>
<td>120,427</td>
<td>125,000</td>
<td>96.34</td>
</tr>
<tr>
<td>53B4</td>
<td>152-800</td>
<td>HiSeq</td>
<td>CTC</td>
<td>15</td>
<td>23,972</td>
<td>1,735</td>
<td>123,510</td>
<td>130,000</td>
<td>95.01</td>
</tr>
<tr>
<td>25P3</td>
<td>173-500</td>
<td>GAIIx</td>
<td>ACG</td>
<td>3</td>
<td>45,211</td>
<td>7,373</td>
<td>92,393</td>
<td>90,000</td>
<td>100*</td>
</tr>
<tr>
<td>3A24</td>
<td>173-500</td>
<td>HiSeq</td>
<td>GGG</td>
<td>7</td>
<td>42,945</td>
<td>2,203</td>
<td>115,648</td>
<td>125,000</td>
<td>92.52</td>
</tr>
<tr>
<td>18L4</td>
<td>AA12-850</td>
<td>GAIIx</td>
<td>GGG</td>
<td>8</td>
<td>35,993</td>
<td>1,897</td>
<td>120,953</td>
<td>150,000</td>
<td>80.64</td>
</tr>
<tr>
<td>54O7</td>
<td>AA12-850</td>
<td>GAIIx</td>
<td>CCC</td>
<td>5</td>
<td>54,064</td>
<td>5,244</td>
<td>110,897</td>
<td>120,000</td>
<td>92.41</td>
</tr>
<tr>
<td>85B18</td>
<td>AA12-850</td>
<td>GAIIx</td>
<td>CCC</td>
<td>4</td>
<td>65,897</td>
<td>5,948</td>
<td>124,801</td>
<td>125,000</td>
<td>99.84</td>
</tr>
<tr>
<td>67D17</td>
<td>AA12-850</td>
<td>GAIIx</td>
<td>TGC</td>
<td>2</td>
<td>51,448</td>
<td>41,556</td>
<td>93,004</td>
<td>90,000</td>
<td>100*</td>
</tr>
<tr>
<td>87L10</td>
<td>AA12-850</td>
<td>GAIIx</td>
<td>CTG</td>
<td>5</td>
<td>45,632</td>
<td>393</td>
<td>90,035</td>
<td>85,000</td>
<td>100*</td>
</tr>
<tr>
<td>60F8</td>
<td>AA12-850</td>
<td>GAIIx</td>
<td>GCT</td>
<td>5</td>
<td>65,438</td>
<td>2,807</td>
<td>94,236</td>
<td>85,000</td>
<td>100*</td>
</tr>
<tr>
<td>70N1</td>
<td>AA12-850</td>
<td>GAIIx</td>
<td>CAC</td>
<td>5</td>
<td>77,983</td>
<td>4,307</td>
<td>117,921</td>
<td>120,000</td>
<td>98.27</td>
</tr>
<tr>
<td>72F19</td>
<td>AA12-850</td>
<td>GAIIx</td>
<td>GGG</td>
<td>1</td>
<td>108,194</td>
<td>0</td>
<td>108,194</td>
<td>100,000</td>
<td>100*</td>
</tr>
<tr>
<td>43F13</td>
<td>W07-375</td>
<td>GAIIx</td>
<td>CGT</td>
<td>8</td>
<td>31,223</td>
<td>2,815</td>
<td>101,433</td>
<td>135,000</td>
<td>75.14</td>
</tr>
<tr>
<td>65G23</td>
<td>W07-375</td>
<td>GAIIx</td>
<td>TGC</td>
<td>4</td>
<td>51,029</td>
<td>708</td>
<td>100,335</td>
<td>110,000</td>
<td>90.80</td>
</tr>
<tr>
<td>66C22</td>
<td>W07-375</td>
<td>GAIIx</td>
<td>CTG</td>
<td>5</td>
<td>40,853</td>
<td>1,013</td>
<td>68,934</td>
<td>90,000</td>
<td>76.59</td>
</tr>
<tr>
<td>68G11</td>
<td>W07-375</td>
<td>GAIIx</td>
<td>ACG</td>
<td>2</td>
<td>59,391</td>
<td>35,367</td>
<td>94,758</td>
<td>100,000</td>
<td>94.76</td>
</tr>
<tr>
<td>69B19</td>
<td>W07-375</td>
<td>GAIIx</td>
<td>CAC</td>
<td>4</td>
<td>45,817</td>
<td>2,609</td>
<td>82,878</td>
<td>90,000</td>
<td>92.09</td>
</tr>
<tr>
<td>85B7</td>
<td>W07-375</td>
<td>GAIIx</td>
<td>CGT</td>
<td>8</td>
<td>44,096</td>
<td>1,536</td>
<td>74,289</td>
<td>115,000</td>
<td>64.60</td>
</tr>
<tr>
<td>67L9</td>
<td>X01-825</td>
<td>GAIIx</td>
<td>GCT</td>
<td>4</td>
<td>69,962</td>
<td>1,743</td>
<td>94,012</td>
<td>100,000</td>
<td>94.01</td>
</tr>
<tr>
<td>96K15</td>
<td>X01-825</td>
<td>GAIIx</td>
<td>AGC</td>
<td>19</td>
<td>70,280</td>
<td>2,898</td>
<td>270,199</td>
<td>225,000</td>
<td>100*</td>
</tr>
<tr>
<td>48L4</td>
<td>H04-850</td>
<td>HiSeq</td>
<td>GGC</td>
<td>11</td>
<td>16,929</td>
<td>1,083</td>
<td>73,471</td>
<td>115,000</td>
<td>63.89</td>
</tr>
<tr>
<td>784A</td>
<td>H04-850</td>
<td>HiSeq</td>
<td>TCG</td>
<td>8</td>
<td>27,094</td>
<td>1,285</td>
<td>95,533</td>
<td>120,000</td>
<td>79.61</td>
</tr>
<tr>
<td>38N24</td>
<td>H04-850</td>
<td>HiSeq</td>
<td>CAC</td>
<td>11</td>
<td>31,170</td>
<td>1,307</td>
<td>98,093</td>
<td>130,000</td>
<td>75.46</td>
</tr>
<tr>
<td>60J6</td>
<td>268-580</td>
<td>HiSeq</td>
<td>CGG</td>
<td>15</td>
<td>30,411</td>
<td>1,672</td>
<td>121,356</td>
<td>150,000</td>
<td>80.90</td>
</tr>
<tr>
<td>54Q21</td>
<td>726-665</td>
<td>GAIIx</td>
<td>AGC</td>
<td>7</td>
<td>52,147</td>
<td>1,951</td>
<td>125,069</td>
<td>125,000</td>
<td>100*</td>
</tr>
<tr>
<td>41B19</td>
<td>726-665</td>
<td>HiSeq</td>
<td>CCC</td>
<td>15</td>
<td>20,958</td>
<td>1,280</td>
<td>106,603</td>
<td>100,000</td>
<td>100*</td>
</tr>
<tr>
<td>62A9</td>
<td>726-665</td>
<td>HiSeq</td>
<td>CCA</td>
<td>13</td>
<td>22,638</td>
<td>651</td>
<td>95,618</td>
<td>95,000</td>
<td>100*</td>
</tr>
</tbody>
</table>
Ab initio Gene Annotation

- Gene prediction carried out using AUGUSTUS
  - Arabidopsis as the gene prediction model
  - RNA-Seq data from ‘Jefferson’ (Data from Mockler lab)
  - Amino acid sequences of predicted genes were BLAST(P) searched for protein homology

<table>
<thead>
<tr>
<th>Contig</th>
<th>RAPD</th>
<th>Contig Source</th>
<th>No. of BACs sequenced</th>
<th>No. of BACs in the contig</th>
<th>No. of genes at four percentages of RNA-Seq transcript support</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>173-500</td>
<td>Resistant</td>
<td>2</td>
<td>2</td>
<td>34    13    8    7</td>
</tr>
<tr>
<td>02</td>
<td>AA12-850</td>
<td>Resistant</td>
<td>7</td>
<td>7</td>
<td>30    11    6    5</td>
</tr>
<tr>
<td>03</td>
<td>AA12-850</td>
<td>Susceptible</td>
<td>3</td>
<td>3</td>
<td>62    16    10   6</td>
</tr>
<tr>
<td>04</td>
<td>W07-375</td>
<td>Resistant</td>
<td>3</td>
<td>3</td>
<td>37    12    6    5</td>
</tr>
<tr>
<td>05</td>
<td>W07-375</td>
<td>Susceptible</td>
<td>2</td>
<td>2</td>
<td>15    4     1    0</td>
</tr>
<tr>
<td>06</td>
<td>X01-825</td>
<td>Resistant</td>
<td>1</td>
<td>1</td>
<td>37    9     6    6</td>
</tr>
<tr>
<td>07</td>
<td>X01-825</td>
<td>Susceptible</td>
<td>1</td>
<td>4</td>
<td>18    8     6    3</td>
</tr>
</tbody>
</table>

Total 233 73 43 32
Potential Candidate Genes (2 of 5)

Predicted gene Contig4_g19
- Belongs to p-loop NTPase superfamily
- Includes NBS-LRR type disease resistance proteins
- NBS-LRR is the major class of R genes so far identified

Predicted gene Contig4_g25
- Belongs to F-box superfamily
- Plant F-box genes – one of the largest multigene superfamilies
- Controls many important biological functions including disease resistance

Sathuvalli et al., in preparation
Incompatibility in Hazelnut

- Sporophytic, one S-locus, 33 alleles
- Dominance or co-dominance in pollen
- Co-dominance in stigmas
- Determined by fluorescence microscopy

If the same allele is expressed by the stigma and the pollen, the cross is incompatible.
Incompatibility Testing using Fluorescence Microscopy

Compatible
• Excellent germination
• Long parallel tubes

Incompatible
• Poor germination
• Short tubes, bulbs

Mehlenbacher, 2011
S-locus Mapping

LG5 Map

Chromosome walking in progress

Fine mapping of S-locus region

Sathuvalli and Mehlenbacher (2011) Genome 54:862-867
Development of SSRs

- 24 SSRs developed from BACs sequenced in the EFB resistance region

- SSRs (3-5 bp repeats) are being developed from the hazelnut transcriptome (Data from Mockler lab)
  - Leaf: 158 SSRs
  - Bark: 129 SSRs
  - Catkins: 81 SSRs

- Screening 368 SSRs with 24 diverse genotypes identified 158 (43%) as polymorphic

- Mapping and characterization studies are underway for these SSRs

Sathuvalli and Mehlenbacher (2012) Tree Genetics & Genomes (Submitted)
Genomic tools are being effectively used in hazelnut breeding

- Map-based cloning identified BACs in the EFB resistance region
  - 5 candidate genes identified

- Map-based cloning of S-locus initiated

- SSR markers were developed in the EFB resistance region

- SSR markers are being developed from transcriptome data
Acknowledgements

David C Smith
Todd C Mockler (Genome Sequencing)
Nahla V Bassil
Cristino Montes
Becky McCluskey
Kahraman Gurcan
Erik Rowley
Brian Knaus (Bioinformatics)
Rich Cronn
Chris Sullivan (CGRB)
Wayne Wood

Funding Sources

Oregon Hazelnut Commission

USDA

Specialty Crops Research Initiative Grant
2009-51181-06028