SNP Diversity in Cultivated Tomato

Joanne A. Labate
Larry D. Robertson
Angela M. Baldo

USDA-ARS Plant Genetic Resources Unit
Geneva, NY
Recent diversity studies of cultivated tomato

fresh market, processing, heirloom lines (Van Deynze et al. 2007)
European commercial greenhouse cultivars (van Berloo et al. 2008)
Italian landraces, vintage, modern cultivars (Mazzucato et al. 2008)
Greek landraces, cultivars (Terzopoulus and Bebeli 2008)
cultivars, hybrids, elite lines (Chen et al. 2009)
landraces from centers of diversity, cultivars (Labate et al. 2009)

• discover polymorphic markers that may be transferable to elite germplasm
• utilize unimproved, cultivated germplasm as sources of new alleles
PGRU Diversity Panels (30-50 NPGS accessions)

- DNA sequences 50 loci (gene fragments)
- SNP genotypes (D. Francis lab)
- Agronomic traits and fruit quality (2008, 2009)
Genetic diversity among 31 tomato lines based on 50 sequenced markers via covariance matrix with data standardization

- primary centers of diversity (n = 14)
- secondary centers of diversity (n = 4)
- countries contiguous with primary centers (n = 12)
- line TA496

Labate et al. (2009) TAG 118:1005–1014
Mean diversity across 50 loci for three *S. lycopersicum* samples

<table>
<thead>
<tr>
<th>Population</th>
<th>no.(^a)</th>
<th>no. of alleles</th>
<th>(F)</th>
<th>(H_e)^b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary center</td>
<td>14</td>
<td>2.3</td>
<td>0.793</td>
<td>0.254</td>
</tr>
<tr>
<td>Contiguous w/ primary center</td>
<td>12</td>
<td>1.7</td>
<td>0.947</td>
<td>0.152</td>
</tr>
<tr>
<td>Secondary centers</td>
<td>6</td>
<td>1.2</td>
<td>1.000</td>
<td>0.092</td>
</tr>
</tbody>
</table>

\(^a\)Number of plants sampled

\(^b\)Expected heterozygosity = \(2n(1-\sum x^2)/(2n-1)\) where \(n\) is the number of sampled chromosomes, \(x\) is allele frequency; each haplotype was treated as an allele
Number of alleles per locus in *S. lycopersicum*, 50 loci
Allele frequency in *S. lycopersicum*, 50 loci

![Graph showing allele frequency distribution with bars for different frequency ranges in a sample.](image-url)
Within-locus linkage disequilibrium in *S. lycopersicum*, 21 loci

Between-locus linkage disequilibrium in *S. lycopersicum*

41 loci, 797 pairwise comparisons:

• 152 pairs (0.191) in LD \((P \leq 0.05)\)
• for 75/152 pairs with known map locations, only 8 pairs (0.107) were located on same chromosome

### Diversity estimates for various crops and wild relatives

<table>
<thead>
<tr>
<th>Species</th>
<th>Mating system&lt;sup&gt;a&lt;/sup&gt;</th>
<th>SNPs/kb&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soybean</td>
<td>S</td>
<td>1.1</td>
<td>Zhu et al. 2003</td>
</tr>
<tr>
<td>Tomato</td>
<td>S</td>
<td>1.3</td>
<td>Labate et al. 2009</td>
</tr>
<tr>
<td>Sorghum</td>
<td>S</td>
<td>2.2</td>
<td>Hamblin et al. 2006</td>
</tr>
<tr>
<td>Asian rice</td>
<td>S</td>
<td>2.3</td>
<td>Caicedo et al. 2007</td>
</tr>
<tr>
<td>Wild barley</td>
<td>S</td>
<td>7.5</td>
<td>Morrell et al. 2005</td>
</tr>
<tr>
<td><em>S. chilense</em></td>
<td>SI</td>
<td>11.0</td>
<td>Arunyawat et al. 2007</td>
</tr>
<tr>
<td><em>S. peruvianum</em></td>
<td>SI</td>
<td>12.9</td>
<td>Arunyawat et al. 2007</td>
</tr>
</tbody>
</table>

<sup>a</sup>S, self-fertilizing; SI, self-incompatible

<sup>b</sup>Expected heterozygosity for a pair of SNPs sampled at random
Haplotype network
C2_At1g32130
(430 bp)

S. arcanum

n = 3

n = 31

A/G

C/T G/T G/T G/C

synonymous

non-syn, conservative

non-syn, nonconservative

noncoding (no shading)

S. arcanum
Haplotype network
2819_5
(656 bp)

“hypothetical introgression”

noncoding (no shading)
- synonymous
- non-syn, conservative
- non-syn, nonconservative

PI 258478 (Peru, 1959)
Mean pairwise divergence between sequences at two control loci

- tom = *S. lycopersicum* “tomate”
- pw = *S. lycopersicum* “Peru Wild”
- pim = *S. pimpinellifolium*
- arc = *S. arcanum*
- pen = *S. pennellii*
- hab = *S. habrochaites*
Mean pairwise divergence between sequences at five loci with hypothetical introgressions

tom = *S. lycopersicum* “tomate”
pw = *S. lycopersicum* “Peru Wild”
pim = *S. pimpinellifolium*
arc = *S. arcanum*
pen = *S. pennellii*
hab = *S. habrochaites*
div = *S. lycopersicum* “diverged allele”
Summary and conclusions

• diversity within *S. lycopersicum* similar to other selfing crops (1 – 2 SNPs/kb)
  • most loci (80%) carry at least two alleles
  • variant alleles are rare in sample (≤ 10%)
  • individuals are highly homozygous ($F = 0.8 – 1.0$)
  • low to moderate population diversity ($H_e = 0.09 – 0.25$)
• high within-locus LD
• moderate to low between-locus LD
• highly diverged haplotypes are observed ~10% (cryptic introgressions?) Islam et al. 2004
Thank you!

Funded by CRIS project 1910-21000-019-00D