The Origin of Orange Pigment in Carrot

Sunny Pei - Jan. 11th, 2015
Domestication Genomics Workshop
International Plant & Animal Genome (IPAG)
San Diego, CA, USA

Overview
- What do we currently know about carrot domestication?
  - Historical and molecular insight
- What questions remain unanswered or incomplete?
  - Is the origin in Central Asia or Asia Minor?
  - What regions of the genome were impacted during domestication?
  - What is the extent of linkage disequilibrium in carrot?
- Preliminary results using geographically well distributed samples from the carrot PI collection to address the above questions
  - Phylogenetic relatedness, population structure, nucleotide diversity and an estimate of LD in carrot

Overview
- What do we currently know about carrot domestication?
  - Historical and molecular insight
- What questions remain unanswered or incomplete?
  - Is the origin in Central Asia or Asia Minor?
  - What regions of the genome were impacted during domestication?
  - What is the extent of linkage disequilibrium in carrot?
- Preliminary results using geographically well distributed samples from the carrot PI collection to address the above questions
  - Phylogenetic relatedness, population structure, nucleotide diversity and an estimate of LD in carrot

Carrot Domestication - historical
- Wild carrot (seed) was first used as a medicinal plant in ancient Rome
- 98-138 AD Roman tavern with a painting of food depicting carrot or parsnip
- Pre-900s purple and yellow carrot varieties in Afghanistan and surrounding vicinity
- 11th century script with orange/yellow carrot
- 1100 AD domesticated carrots moved into SW Europe
- 13th century carrots moved into the Eastern parts of Asia
- European cultivated carrots found in Americas soon after Columbus' first visit
- No mention of orange
- 1600s orange colored carrots frequently described

Banga, 1963

Carrot Domestication - molecular
- Wild vs. Domesticated separation (Shim and Jorgensen, 2000, Theor Appl Genet)
- Eastern vs. Western separation (Clotault et al., 2010, Theor Appl Genet; Baranski et al., 2012, Genet Resour Crop Evol; Grzebelus et al., 2013, Mol Breed, Rong et al., 2014 BMC Genomics)
- Wild carrots from Central Asia are the closest genetic relatives to domesticated carrots (Iorizzo et al., 2013, Am J Bot)
- Limited sample size
- Domesticated carrot maintains a high level of genetic diversity

Remaining questions
- Will a more robust analysis of population structure and phylogeny still point to Central Asia as the domestication center of carrot?
- What regions of the genome were impacted during domestication?
- What is the extent of linkage disequilibrium in carrot?

Remaining questions
- Will a more robust analysis of population structure and phylogeny still point to Central Asia as the domestication center of carrot?
- What regions of the genome were impacted during domestication?
- What is the extent of linkage disequilibrium in carrot?

Here we present our preliminary results analyzing 333 geographically distributed wild and domesticated PIs
Geographic distribution of materials

**Open Pollinated Material**

<table>
<thead>
<tr>
<th>Geographic region</th>
<th>Country</th>
<th>Number of accessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>South America</td>
<td>Chile, Argentina</td>
<td>5</td>
</tr>
<tr>
<td>North America</td>
<td>United States of America</td>
<td>14</td>
</tr>
<tr>
<td>Europe</td>
<td>Belgium, Bulgaria, Denmark, France, Germany, Hungary, Netherlands, Spain, Sweden, Russia</td>
<td>40</td>
</tr>
<tr>
<td>North Africa</td>
<td>Tunisia</td>
<td>10</td>
</tr>
<tr>
<td>Middle East</td>
<td>Turkey, Syria</td>
<td>20</td>
</tr>
<tr>
<td>Central Asia</td>
<td>Pakistan</td>
<td>13</td>
</tr>
<tr>
<td>South Asia</td>
<td>India</td>
<td>6</td>
</tr>
<tr>
<td>North Asia</td>
<td>China, Japan, Egypt, Georgia, Ethiopia, South Korea, Mongolia, Nepal</td>
<td>12</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>193</td>
</tr>
</tbody>
</table>

**Wild Plant Material**

<table>
<thead>
<tr>
<th>Geographic region</th>
<th>Country</th>
<th>Number of accessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>South America</td>
<td>Chile, Argentina</td>
<td>1</td>
</tr>
<tr>
<td>North America</td>
<td>United States of America</td>
<td>28</td>
</tr>
<tr>
<td>Europe</td>
<td>Czech Republic, France, Germany, Greece, Hungary, Italy, Netherlands, Portugal, Switzerland, Austria, Albania, Poland, Bulgaria</td>
<td>82</td>
</tr>
<tr>
<td>North Africa</td>
<td>Tunisia</td>
<td>10</td>
</tr>
<tr>
<td>Middle East</td>
<td>Turkey, Syria</td>
<td>41</td>
</tr>
<tr>
<td>Central Asia</td>
<td>Uzbekistan, Pakistan</td>
<td>13</td>
</tr>
<tr>
<td>South Asia</td>
<td>China</td>
<td>1</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>180</td>
</tr>
</tbody>
</table>

Genotyping

- Extract DNA from PIs, prepare GBS libraries and sequence
- Run the Tassel 4.0 GBS Pipeline and call SNPs
- After filtering: 284 individuals, 37,864 SNPs (10% missing data, 5% MAF)

Population Structure

- **FastSTRUCTURE**
  - Uses a variational Bayesian framework to approximate inference of the model used by STRUCTURE
  - K range given by model complexity that maximizes marginal likelihood and model components
  - K = 9

Phylogenetic analysis

- **NJ Tree**
  - Clear separation of all groups found with K=9 except wild Asia minor/C. Asia and dom Asia minor/C. Asia
  - Defined split between Dom vs Wild and Eastern vs Western

Population Structure - FastSTRUCTURE

- Uses a variational Bayesian framework to approximate inference of the model used by STRUCTURE
- K range given by model complexity that maximizes marginal likelihood and model components

Substructure within wild populations
Dissemination of domesticated types

Carrot Domestication Traits

- Root Color
- Reduction of lateral branches
- Reduction of bolting
- Enlarged taproot
- Sugar accumulation
- Flavor – terpenoid biosynthesis

What regions of the carrot genome control these traits?

Nucleotide Diversity

<table>
<thead>
<tr>
<th>Chr1</th>
<th>Chr2</th>
<th>Chr3</th>
<th>Chr4</th>
<th>Chr5</th>
<th>Chr6</th>
<th>Chr7</th>
<th>Chr8</th>
<th>Chr9</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild</td>
<td>0.35</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>Dom</td>
<td>0.31</td>
<td>0.30</td>
<td>0.31</td>
<td>0.32</td>
<td>0.32</td>
<td>0.31</td>
<td>0.30</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td>OrDom</td>
<td>0.32</td>
<td>0.28</td>
<td>0.29</td>
<td>0.31</td>
<td>0.28</td>
<td>0.29</td>
<td>0.32</td>
<td>0.29</td>
<td>0.29</td>
</tr>
</tbody>
</table>

100kb sliding window

<table>
<thead>
<tr>
<th>Chr1</th>
<th>Chr2</th>
<th>Chr3</th>
<th>Chr4</th>
<th>Chr5</th>
<th>Chr6</th>
<th>Chr7</th>
<th>Chr8</th>
<th>Chr9</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild</td>
<td>0.40</td>
<td>0.40</td>
<td>0.38</td>
<td>0.39</td>
<td>0.39</td>
<td>0.37</td>
<td>0.37</td>
<td>0.37</td>
<td>0.37</td>
</tr>
<tr>
<td>Dom</td>
<td>0.21</td>
<td>0.17</td>
<td>0.18</td>
<td>0.13</td>
<td>0.21</td>
<td>0.22</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>OrDom</td>
<td>0.14</td>
<td>0.13</td>
<td>0.07</td>
<td>0.10</td>
<td>0.18</td>
<td>0.14</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
</tr>
</tbody>
</table>

LYW/ LONESOME HIGHWAY

inv/Dvi5  chaperone protein dreau (Ox, maim)  SUCROSE PHOSPHATE SYNTHASE  UU5, UUTEN DEFICIENT S  y,III  CCX4, NCE24
Decreased nucleotide diversity at the \( y^2 \) locus

**Linkage Disequilibrium**

- LD typically decays rapidly within outcrossing species
  - 1–5 kb in maize, 1.1 kb in cultivated sunflower, and 300 bp in wild grapevine
- And slowly in selfing species
  - 200 kb in Arabidopsis, 212 kb in barley cultivars, 100–200 kb in rice diverse lines and 250 kb in cultivated soybean
- Carrot is an highly outcrossing species with very high levels of genetic diversity
  - LD Decay ~3.5–9 kb

<table>
<thead>
<tr>
<th>Chr</th>
<th>Chr2</th>
<th>Chr3</th>
<th>Chr4</th>
<th>Chr5</th>
<th>Chr6</th>
<th>Chr7</th>
<th>Chr8</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>7.8384</td>
<td>10.056</td>
<td>8.5749</td>
<td>6.0348</td>
<td>11.1596</td>
<td>17.9306</td>
<td>4.6265</td>
<td>11.4543</td>
</tr>
</tbody>
</table>

**What’s next?**

- More robust tree and population structure analyses
- Modeltest to find the best parameters to construct phylogeny
- STRUCTURE vs FastSTRUCTURE
- Explore regions underneath regions of low nucleotide diversity
- Utilize candidate genes for domestication traits (\( y, y^2 \)) to explore when and where traits were selected
- Include inbred lines and look at various population genomic parameters including genetic diversity, \( F_s \) and LD

**Acknowledgements**

- Simon lab members
  - Dr. Philipp Simon
  - Dr. Massimo Iorrizo
  - Dr. Doug Senalik
  - Mr. Rob Kane
- Support
  - NSF Award #1202666
  - USDA-ARS
- Thank you!