Genotyping-by-Sequencing Reveals the Diversity of the USDA Pisum Diversity Collection

Bill Holdsworth1, Yu Ma2, Peng Cheng2, Rebecca McGee2,3, Clarice Coyne2,3, Jim Myers4, Michael Gore1, and Michael Mazourek1

1Cornell University, 2Washington State University, 3USDA-ARS, 4Oregon State University

 USDA Pisum Diversity Collection

Collection of 477 accessions from 61 countries. USDA PSP core collection + P. fulvum + Asian accessions + Cultivars & Breeding Lines

Genotyping-by-Sequencing Overview

Restriction Enzyme Considerations: Read Depth vs. Genomic Coverage

Gene Enrichment

Differential SNP Identification using UNEAK vs. STACKS

Sample Call Rate = 0.5, MAF = .05

Pisum sativum

17,989 SNPs

Pisum fulvum

27,944 SNPs

Avg. Read Depth

P. sativum

P. fulvum

% Missing Data

P. sativum

P. fulvum

Generating SNPs

All Accessions – Sample Call Rate 0.2, MAF 0.01

<table>
<thead>
<tr>
<th></th>
<th>UNEAK</th>
<th>STACKS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total SNPs</td>
<td>61,386</td>
<td>47,947</td>
</tr>
<tr>
<td>% Missing Data</td>
<td>45.8</td>
<td>46</td>
</tr>
</tbody>
</table>

Sample Call Rate 0.2, MAF .01

<table>
<thead>
<tr>
<th></th>
<th>Pisum fulvum</th>
<th>Pisum sativum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total SNPs</td>
<td>43,861</td>
<td>59,653</td>
</tr>
<tr>
<td>% Missing Data</td>
<td>20-23%</td>
<td>20-26%</td>
</tr>
</tbody>
</table>

45K SNPs is, on average, 1 SNP /100 kb of P. sativum genome, but is enough for trait mapping considering large linkage disequilibrium in pea....

Linkage Disequilibrium in Pea

Map positions used from ~1300 markers from ‘Kiflica’ x ‘Aragorn’ RIL map

LD decays to background genome levels at ~8.5 cM
Association mapping – “A” Gene

“A” phenotype

“a” phenotype

Principal Components Analysis – *P. spp.*

Principal Components Analysis – *Pisum sativum*

PCA Defines Biogeography - *Pisum sativum*

Structure Among *P. sativum*
Principal Components Analysis - *P. sativum*

Collected Accessions by Subpopulation Groupings

Collected Accessions by Subpopulation Groupings

Collected Accessions by Subpopulation Groupings

Large Divergence Between Subpopulations Fst

<table>
<thead>
<tr>
<th>Fst Estimates Between Subpopulation Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Western&quot;</td>
</tr>
<tr>
<td>&quot;Western&quot;</td>
</tr>
<tr>
<td>Admixed</td>
</tr>
<tr>
<td>&quot;Eastern&quot;</td>
</tr>
<tr>
<td><em>P. fulvum</em></td>
</tr>
</tbody>
</table>
Few Accessions Sufficient for Capturing Diversity in Pea

### Defining a Core Subset from All Accessions (477)

<table>
<thead>
<tr>
<th>Filtered SNP subsets</th>
<th>~3500 SNPs</th>
<th>~13,300 SNPs</th>
<th>~61,400 SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Allele Capture</td>
<td>36 (2 P. fulvum)</td>
<td>64 (3 P. fulvum)</td>
<td>327 (21 P. fulvum)</td>
</tr>
<tr>
<td>Diversity Index</td>
<td>10</td>
<td>15</td>
<td>19</td>
</tr>
</tbody>
</table>

Diversity Index maximized weighted Modified Roger’s Genetic Distance and Shannon’s Diversity Index

### Conclusions

- GBS with ApeKI effective method to generate SNPs *de novo* in *Pisum* spp.
- Utilization of multiple “non-reference” pipelines can increase total number of SNPs
- Development of large SNP dataset useful for mapping traits in *Pisum*, especially in light of large LD blocks
- Additional diversity in *Pisum fulvum*
- Population substructure driven by dispersal across trade routes?
- Alleles largely conserved across germplasm pools
- Identification of germplasm pools potentially useful to breeders

### Acknowledgements

- Graduate Special Committee
  - Michael Mazourek
  - Mike Gore
  - Chris Smart
  - Susan McCouch
- Mazurek Lab Group
  - Elissa Gazave
  - Alex Lipka
- Cornell Computational Biology Service Unit
- Robert Bukowski
- Cornell Institute for Genomic Diversity
  - Sharon Mitchell
  - Charlotte Acharya
- Funding Sources
  - National Plant Germplasm System-Pisum Crop Germplasm Committee Grant
  - USDA National Institute of Food and Agriculture, Plant Breeding and Education Grant #2010-85117-20551
- Graduate Special Committee
  - Michael Mazourek
  - Mike Gore
  - Chris Smart
  - Susan McCouch
- Mazourek Lab Group
  - Elissa Gazave
  - Alex Lipka

### UNEAK Informatics Pipeline


http://www.plosgenetics.org/article/info:doi/10.1371/journal.pgen.1003215