Genome-Wide View of Breeding History and Selection in North American Maize

Dr. Jeffrey Ross-Ibarra
University of California at Davis

Patterns of diversity across the genome are reflective of the historical processes of selection, drift, and mutation that have acted on a species during its evolution. Here, we apply population genetic analysis of two genomic datasets to assess the spatial and temporal signatures of evolution in the maize genome. We first describe genome-wide patterns of evolution during two epochs of selection, domestication and modern breeding, using full-genome resequencing of a number of maize and teosinte lines. Then, using a chronologically-stratified panel of corn belt lines, we dissect temporal patterns of ancestry and selection in greater detail across the 80+ years of North American corn breeding. We find distinct impacts of selection during domestication and modern breeding on diversity and gene expression, identify candidate regions targeted by selection, and describe changes in genetic structure and ancestry during the evolution of modern corn belt lines.
Acknowledgements

Ed Buckler (Cornell, USDA)
Jeff Glaubitz (Cornell)
Major Goodman (NC State)
Mike McMullen (Missouri, USDA)
Chi Song (BGI)
Nathan Springer (Minnesota)
Doreen Ware (CSHL, USDA)
Xun Xu (BGI)
HISTORICAL GENOMICS OF MAIZE

Spatial resolution -- genome resequencing

Temporal resolution -- SNP genotyping
HISTORICAL GENOMICS OF MAIZE

Spatial resolution -- genome resequencing

Temporal resolution -- SNP genotyping
Maize Hapmap I: low-copy fraction of genome

- Re-sequenced low-copy portion of genome of 27 inbreds
- High diversity genome, but large regions of low recombination
- Identified low diversity putative targets of selection
- No comparison to older material
- Cannot distinguish domestication from modern improvement

Gore et al. 2009 Science
Maize Hapmap II: full genome resequencing

<table>
<thead>
<tr>
<th>Sequenced Genomes</th>
<th>Sequence Depth</th>
<th>Locality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved</td>
<td>35</td>
<td>5.04 X Temperate, 13 Tropical, 3 Mixed, 6 Chinese</td>
</tr>
<tr>
<td>Landraces</td>
<td>23</td>
<td>5.34 X South American, 14 North American</td>
</tr>
<tr>
<td>parviglumis</td>
<td>14</td>
<td>3.81 X 4 Jalisco/Nayarit, 9 Balsas, 1 Oaxaca, 1 Oaxaca</td>
</tr>
<tr>
<td>mexicana</td>
<td>2</td>
<td>6.83 X 1 Jalisco, 1 Morelos</td>
</tr>
<tr>
<td>Tripsacum</td>
<td>1</td>
<td>12.62 X Colombia</td>
</tr>
</tbody>
</table>

75 resequenced genomes & >21M SNPs

Chia et al. Submitted
Hufford et al. Submitted
Recovery from the domestication bottleneck

Genic Tajima's D

- Maize
- Parviglumis

Genic -- few rare variants
(27% SNPs unique)

Hufford et al. Submitted
Recovery from the domestication bottleneck

Genome-wide Tajima’s D

Density

Genic Tajima’s D

Density

recovery from bottleneck

Genome -- excess rare variants
(40% SNPs unique)

Genic -- few rare variants
(27% SNPs unique)

Purifying selection on genes retarded recovery from bottleneck

Hufford et al. Submitted
Scan for selection -- XPCLR

- Cross-Population Composite Likelihood Ratio
- Identifies extended regions of differentiation
- Compares to simulated model
- Define selected regions, estimate $s$

Chen et al. 2010
Genome Research

Reference population
Object population

Ellison et al. 2011 PNAS

COG3-like  FMP25-like
"frq  PAC10-like  SEC14-like  NSL1-like  NCU02261

B

LA  Carb  Out
Artificial selection in the maize genome

- Stronger selection during domestication ($s \approx 0.015$ vs. $s \approx 0.003$)
- Selection strength consistent with archaeology & natural selection
- Continued selection on domestication genes (18% overlap)
- ~3,000 genes in selected regions. Identified ~1100 candidates
- 5-10% of selected regions had no genes

Hufford et al. Submitted
Population-specific selection

![Graph showing population-specific selection](image)

Gore et al. Science 2009
Population-specific selection

- Many population-specific loci
- Stronger overall estimate of selection
- Still weaker than domestication
- Lower power to rule out drift
Candidate genes for maize improvement

Hufford et al. Submitted
### Selection on the transcriptome

<table>
<thead>
<tr>
<th>Candidates</th>
<th>DOM</th>
<th>IMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in expression</td>
<td>*</td>
<td>--</td>
</tr>
<tr>
<td>Decreased variance</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Tissue specificity#</td>
<td>--</td>
<td>overall higher expression</td>
</tr>
<tr>
<td>Dominance in crosses#</td>
<td>--</td>
<td>inter &gt; intra</td>
</tr>
</tbody>
</table>

![Graph showing RMA-normalized expression with candidates and non-candidates contrasted.]

# Data from Stupar et al. 2008 BMC Plant Bio; Sekhon et al. 2011 Plant J.

- Expression at >18K genes of 27 maize and teosinte using Nimblegen array
- Domestication directly selected on candidate gene expression
- Breeding has worked with highly expressed genes
- Modern breeding selected for heterosis in expression
Regions with no genes likely regulatory

Swanson-Wagner et al. Submitted
Genetic load in modern maize

- ~1% of SNPs disrupts predicted protein
- 10,117 premature stop-codons
- 2557 (or 12%) high-confidence genes has a stop-codon in ≥ 1 line
- 870 genes have a segregating stop-codon in ≥ 2 of improved lines
Rapid Centromeric Change

chr5

chr1
Novel variation in teosinte

- 50Mb inversion in teosinte absent in maize
- Found segregating in nearly all teosinte populations
- Associated with altitude, temp, culm diameter

Fang et al. Submitted
Artificial selection across the maize genome

- New mutations and recovery (esp. noncoding) post-domestication
- Weaker selection during improvement
- Different domestication & improvement effects on expression
- Novel variation present in teosinte
- No evidence for an important role of new genes or CNVs
- Future work: ongoing projects to investigate genetic load
HISTORICAL GENOMICS OF MAIZE

Spatial resolution -- genome resequencing

Temporal resolution -- SNP genotyping
Detailed historical look at selection during modern breeding

- Larger sample size
- Increased chronological resolution
- Take into account heterotic group structure
- Focus on corn belt varieties

Resolve caveats of resequencing approach
Genotyping and Sampling

0 99 land races

1 94 <1950 inbreds (Oh43, W22, B14)

2 70 1960-1970 inbreds (B73, 207, Mo17)

3 137 ex-PVP

- 400 N. American corn belt lines
- Genotype on public Illumina 55K
- Tracked population structure, ancestry, and selection over time

van Heerwaarden et al. Submitted
Population structure through time

van Heerwaarden et al. Submitted
Population structure through time
Population structure through time
Population structure through time
Population structure through time

van Heerwaarden et al. Submitted
Population structure through time

van Heerwaarden et al. Submitted
Number and diversity of ancestors decreases through time

van Heerwaarden et al. Submitted
Ancestral contributions of individual lines
Weak landrace structure still evident in modern inbreds
Selection and frequency over time

advantageous alleles show continual increase in frequency over time
Time as a phenotype to identify selection

0

• Identify genetic structure across all time periods (PCA, Tracy-Widom)

1

• Treat time period as a phenotype for association mapping

2

• For each SNP, compare two models:
  • M0: Frequency determined by drift and population structure (covariance matrix of all SNPs)
  • M1: Frequency determined by drift, population structure AND environment (time)

3

• Bayes Factor for each SNP

Coop et al. 2010 Genetics
Temporal association identifies candidates of selection

- 236 Candidate regions (1021 genes) with consistent BF across runs
- Include stress response, lignin biosynthesis, auxin response
- Two of the top candidates ZmErecta (US7847158) and ZmArgos (US7834240) patented for plant growth, yield
Temporal association identifies candidates of selection

- 236 Candidate regions (1021 genes) with consistent BF across runs
- Include stress response, lignin biosynthesis, auxin response
- Two of the top candidates ZmErecta (US7847158) and ZmArgos (US7834240) patented for plant growth, yield
Selective sweeps

[Diagram showing the concept of selective sweeps in genetics. The left side of the diagram illustrates a sequence of DNA segments with varying levels of diversity. The right side shows a graph of diversity over a genome sequence, with a highlighted region indicating a selective sweep.]
No abnormal haplotypes, ancestry at selected sites
No abnormal haplotypes, ancestry at selected sites
Genome-wide ancestry not determined by selection
Weak correlation between good alleles and contribution to modern lines
Weak correlation between good alleles and contribution to modern lines
Structure and selection in corn belt maize

- Increasing structure, LD, less diverse ancestry over time
- Structure evident in landraces as is contribution to modern lines
- Selected loci can be identified controlling for population structure
- Selected loci show few signs of hard sweeps
- No strong correlation between genotype at beneficial alleles and contribution
- Suggests selection on mostly quantitative traits of small effect
Some final thoughts on adaptation...
Fisher-Orr Model of Adaptation

Modern corn

Phenotypic space
FISHER-ORR MODEL OF ADAPTATION

modern corn

phenotypic space

teosinte
Fisher-Orr model of adaptation

Phenotypic space

Modern corn • Teosinte

Diagram showing the phenotypic space with modern corn and teosinte points.
FISHER-ORR MODEL OF ADAPTATION

landraces

phenotypic space
Fisher-Orr model of adaptation

Landraces

Phenotypic space
- Effect sizes of ear vs. other consistent with Fisher-Orr model
- Larger effects $\rightarrow$ larger fitness difference $\rightarrow$ stronger selection

Brown et al. 2011 PLoS Genetics
Thank You!