Applications of Genotyping-by-Seqencing for Wheat Breeding and Genetics

Jesse Poland
Kansas State University

Next Generation Genomics and Integrated Breeding for Crop Improvement
4th International Workshop
ICRISAT, India
Feb 19, 2014
The Breeding Funnel

Years

Crossing

Early Generation Testing

Prelim Yield testing (thousands)

Replicated Yield testing (hundreds)

Advanced Yield testing (tens)

Varieties (one)
The Breeding Funnel

- Early Generation Testing (thousands)
- Preliminary Yield testing (hundreds)
- Replicated Yield testing (hundreds)
- Advanced Yield testing (tens)
- Genomic Selection
- Back cross conversion
- Parent selection
- F₂ enrichment (MAS)

Varieties (one)
Molecular Markers and Objectives

Single Locus Typing
- Target known genes
- Few loci (<10)

✅ Marker assisted selection
✅ Backcross conversion

Whole-genome Profile
- Assay whole genome
- Many loci (thousands)

✅ Genomic Selection (AM)
✅ Background selection
✅ Diversity study
✅ Germplasm typing

Cost per data point

Cost per sample
Genotyping-by-sequencing (GBS)

Why use sequencing for genotyping rather than array based methods?
+ Amazing developments in sequencing output
+ Very good for wheat where polyploidy and duplications cause problems with hybridization/PCR assays
+ Polymorphism discovery simultaneous with genotyping
+ No ascertainment bias
+ Low per sample cost

- Complex bioinformatics
  - Requires paradigm shift in molecular markers

---

Cost per Sequenced Mb

Aug 22, 2013
Genotyping-by-sequencing (GBS)

“…massively parallel sequencing of multiplexed reduced-representation genomic libraries.”

“massively parallel sequencing” = sequencing on Illumina platform

“multiplex” = using DNA barcode (unique 5-10bp)
   - unique DNA sequence synthesized on the adapter
   - pool 48-384 samples together

“reduced-representation” = use restriction enzyme to capture only the portion of the genome flanking restriction sites
   - methylation-sensitive restriction enzymes
   - Target specific (rare, low-copy) sites in genome
   - PstI (CTGCAG), MspI (CCGG)
Application of GBS:

Genomic Selection
Genomic Selection

Needed:
1) Training Population (genotypes + phenotypes)
2) Selection Candidates (genotypes)

- Accurate phenotypes
- Inexpensive, high-density genotypes


Aug 22, 2013
Using GBS for GS

Is genotyping-by-sequencing a suitable marker platform for genomic selection?

CIMMYT Semi-Arid Wheat Screening Nursery (SAWSN)

- N = 254, advanced lines
- Replicated field trials, Cd. Obregon, Mexico

Using GBS for GS

CIMMYT Semi-Arid Wheat Screening Nursery (SAWSN)

- **GBS**: *PstI-MspI*, 96-plex
  - HiSeq2000 = 180M – 210M reads / lane
  - 41,371 SNPs → 35K
  - DArT markers (n = 1,729)
- **Ridge-regression (rrBLUP)**

---


Prediction of wheat quality

CIMMYT elite breeding lines (n=164)
Cycle 45 International Bread Wheat Screening Nursery (C45IBWSN)

Replicated yield tests
✓ 2009 & 2010
✓ 6 environments
One replication for quality testing
✓ milling
✓ dough rheology
✓ baking tests
Best Linear Unbiased Estimate (BLUE)

Cross-validation (x100)
✓ Training sets of n=134
✓ Validation sets of n=30
- thousand kernel weight
- mix time
- pup loaf volume

Genotyping-by-sequencing
15,330 SNPs (imputed with MVN-EM)(rrBLUP)

Sarah Battenfield, KSU
Prediction of wheat quality

<table>
<thead>
<tr>
<th>Training Population</th>
<th>Cross valid both years</th>
<th>Cross valid both years</th>
<th>Cross valid both years</th>
<th>2011</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Size (n)</td>
<td>1138</td>
<td>995</td>
<td>712</td>
<td>995</td>
<td>712</td>
</tr>
<tr>
<td>Testing Population</td>
<td>Cross valid both years</td>
<td>Cross valid both years</td>
<td>Cross valid both years</td>
<td>2010</td>
<td>2011</td>
</tr>
<tr>
<td>Testing Size (n)</td>
<td>569</td>
<td>712</td>
<td>995</td>
<td>712</td>
<td>995</td>
</tr>
</tbody>
</table>

Prediction Accuracies ($r$)

- **Test Weight**: 0.725***, 0.723***, 0.715***, 0.312***, 0.192***
- **Grain Hardness**: 0.513***, 0.510***, 0.495***, 0.005, 0.056
- **Grain Protein**: 0.630***, 0.629***, 0.620***, 0.400***, 0.335***
- **Flour Protein**: 0.604***, 0.602***, 0.589***, 0.394***, 0.284***
- **Flour SDS Index**: 0.666***, 0.666***, 0.661***, 0.433***, 0.461***
- **Mixograph Mix Time**: 0.718***, 0.715***, 0.707***, 0.535***, 0.499***
- **Alveograph W**: 0.697***, 0.695***, 0.683***, 0.512***, 0.475***
- **Alveograph P/L**: 0.476***, 0.474***, 0.466***, 0.323***, 0.278***
- **Loaf Volume**: 0.638***, 0.634***, 0.625***, 0.358***, 0.333***

Sarah Battenfield, KSU

Aug 22, 2013
Regional Performance Nursery

Established 1931
✓ 1992 – Present
✓ 39 Locations
✓ 80,000 Plots
✓ 350,000 Obs.
792 SRPN + 428 NRPN Entries
✓ 44,924 SNPs
✓ 3,966 SNPs > 80%

Trevor Rife, KSU
Genomic Selection Accuracy: SRPN
RPN Locations: Evaluating Environments

Trevor Rife, KSU
Feed the Future Innovation Lab for Applied Wheat Genomics

www.wheatgenetics.org/research/innovation-lab
Application of GBS:

Characterizing genetic diversity
Characterizing genetic diversity

Hybridization Events Forming Modern Wheat

*Triticum urartu* (2n=2x=14, AA)

*~Aegilops speltoides* (2n=2x=14, SS)

*Triticum turgidum* (2n=4x=28, AABB)

*Aegilops tauschii* (2n=2x=14, DD)

*Triticum aestivum* (2n=6x=42, AABBDD)

---

Eric Olson, Michigan State University

Aug 22, 2013
Wheat Genetic Resource Center: 
Aegilops tauschii collection

• 531 unique accessions
• Physiological classifications
• Genotyped with GBS
Genetic groups contrasting to morphological characteristics
WGRC Ae. *tauschii* Collection

**Genetic separation of geographically separated groups**
Genetic Diversity: Wheat vs *Ae. tauschii*

Limited diversity in elite breeding pool
Application of GBS:

Genetic Mapping
Mapping Resistance in Synthetic Populations

Stem rust resistance to race TRTTF and QTHJC in the SynOpDH population


Sandra Dunckel, KSU; Eric Olson, MSU; Matthew Rouse, USDA-ARS CDL
High-density Genetic Map: *Thinopyrum intermedium*

“F$_2$” population

High-density genetic maps for any species

Th. intermedium GISH

pAs1:green
GAA:red

Sr$_44$

T7DL•7J#1S
Application of GBS:

Marker Assisted Selection
“Spiked GBS”
A unified open platform for single marker genotyping and whole-genome profiling

Utilize 1% of sequencing lane for targeted amplicon sequencing

- Cost effective
- High-throughput
- Flexible: Single set of barcodes combined with locus specific primers
WHOLE GENOME PROFILE (GBS) ($10 – 20 PER SAMPLE\textsuperscript{1})

QC & QUANTIFY
NORMALIZE DNA
DIGEST
LIGATE ADAPTERS

POOL SAMPLES
PCR AMPLIFY
QC & QUANTIFY

“SPIKE” AMPLICON LIBRARY AT 1%

NEXT-GEN SEQUENCING

RAW SEQUENCING DATA
~200M READS

GBS BIOINFORMATICS PIPELINE
~198M READS
~50,000 MARKERS ON 96 INDIVIDUALS 0.5X COVERAGE

TARGETED AMPLICON BIOINFORMATICS PIPELINE
~2M READS
~10 MARKERS ON 384 INDIVIDUALS 500X COVERAGE

SINGLE LOCUS GENOTYPING (TARGET AMPLICONS) (~ $0.03 PER GENOTYPE\textsuperscript{2})

PCR AMPLIFY TARGETS WITH M13 BARCODE PRIMERS

POOL
QC AND QUANTIFY

\textsuperscript{1}THE ESTIMATED COST PER SAMPLE IS BASED ON THE NUMBER OF SAMPLES THAT ARE MULTIPLEXED INTO A SINGLE SEQUENCING RUN AND THE COST OF THE SEQUENCING. PER SAMPLE COST OF $10 CORRESPONDS TO GENOTYPING 190 INDIVIDUALS IN A MULTIPLEX SEQUENCING RUN.

\textsuperscript{2}ESTIMATED COST PER DATA POINT FOR GENOTYPING 10 MARKERS ON 384 INDIVIDUALS.

Aug 22, 2013
“Spiked GBS”: SNP genotyping

- 96 winter wheat accessions
- GBS library
- Amplify 4 SNP loci and add at 1%
- “Converted” KASPar Markers
  (removed selective bp, add tail for barcode)
Application of GBS:

Variety Identification and Typing
Variety Confirmation and Identification

Breeder Seed → Foundation Seed → Production

Confirm Variety → Identify Variety
Mixed up seed? A tail of two samples...

- 2 sub-samples from each lot
- Extracted DNA
- Genotyped (along with a larger panel)
- GBS
- 47,076 DNA markers

< 5% heterozygous markers
- Pure line varieties
- 74% identical markers
- Different varieties

Compared to reference panel and varieties identified as Fuller (#1) and Endurance (#2)