




Genomic technologies and resources for wheat genetics and breeding

Eduard Akhunov

Throckmorton Plant Sciences Center
Kansas State University





With more than **629 million** tons produced annually worldwide, wheat is one of the most important agricultural commodities supplying **40%** of the world's food and **25%** of calories consumed in developing countries.

To meet global food demand by 2020 wheat production should be increased by about **40%**

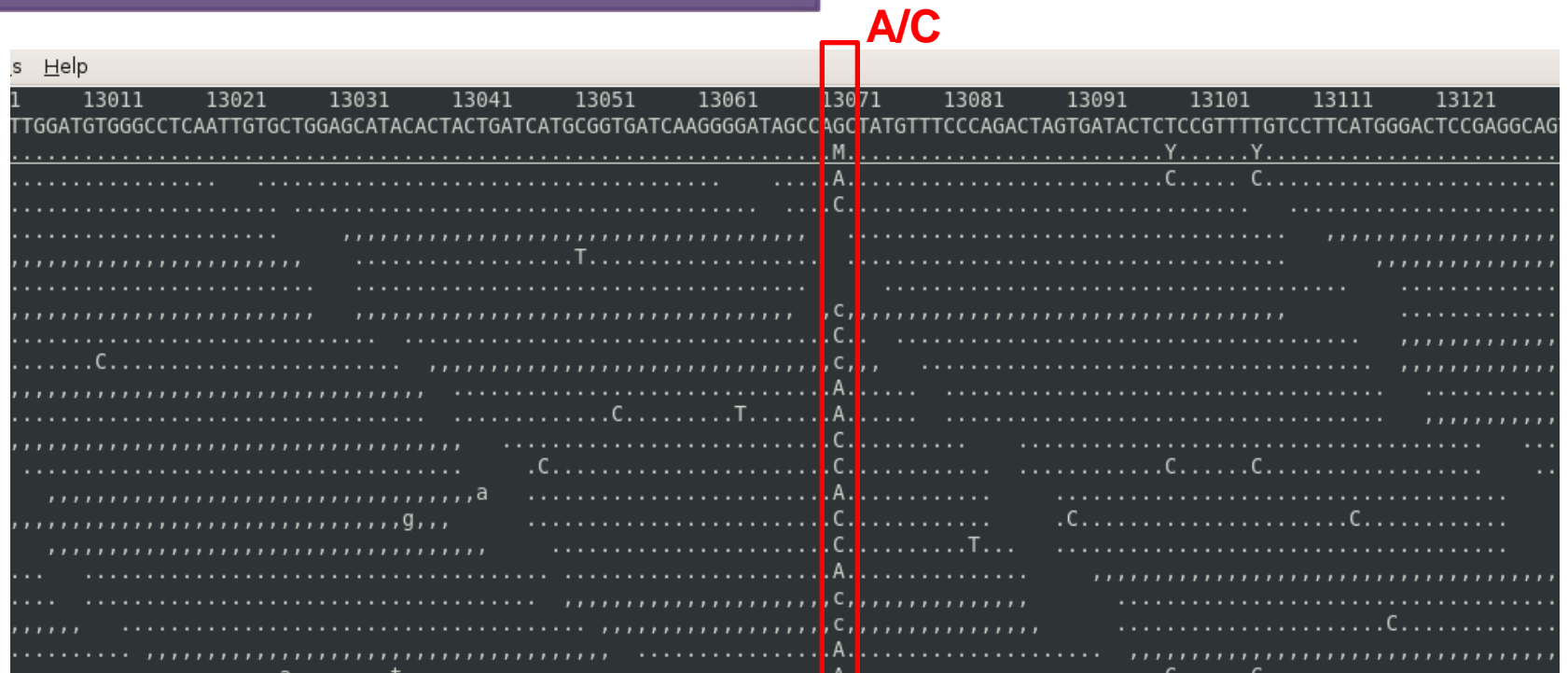
How can we achieve this goal?

- Utilize genetic diversity of wheat and its wild ancestors
- Develop genomic resources and tools
 - High-throughput markers
 - Genome sequence
- Develop high-throughput phenotyping methods
- Use broadly new approaches for gene mapping
 - Association mapping
 - Nested-association mapping
- Adopt new breeding methodologies
 - Marker assisted selection
 - Genomic selection

How can we achieve this goal?

- Utilize genetic diversity of wheat and its wild ancestors
- Develop genomic resources and tools
 - High-throughput markers
 - Genome sequence
- Develop high-throughput phenotyping methods
- Use broadly new approaches for gene mapping
 - Association mapping
 - Nested-association mapping
- Adopt new breeding methodologies
 - Marker assisted selection
 - Genomic selection

Single Nucleotide Polymorphisms



- Widely distributed across genome
- Flexible high-throughput assays
- Cost-effective discovery
- Bi-allelic
- Ascertainment bias
- Validation is required

Single Nucleotide Polymorphisms

Standard strategy

Step 1: Assemble discovery panel (set of diverse lines, 8-32)

Step 2: Sequencing and SNP discovery

Step 3: Development of genotyping assays & validation

Step 4: Genotype large sets of lines (mapping populations, diversity panels)

Wheat SNP database (NSF, 2003-2006, J. Dvorak)

(<http://wheat.pw.usda.gov/SNP/new/index.shtml>)

snpdb/ Haplotype Polymorphism in Polyploid Wheats and the

[Search](#) | [Browse by bin](#) | [Sequences](#) | [Stats](#) | [FAQs](#) | [Disclaimer](#) | [Project home](#)

2000 gene fragments with SNPs 1000 SNPs mapped PCR primers for amplification of gene fragments



ation Grant

Search

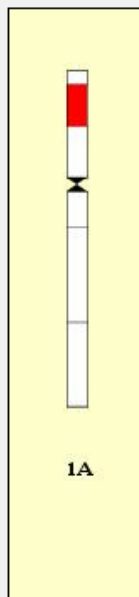
EST accession (optional)
Contig name (optional)
Chromosome/Genome
Lab

Gene Function

EST accession (optional)
Primer Plate Id (optional)
Plant Line
Lab

Comments and questions to: [Database curator](#)

Polymorphism Target Region Mapped to BE399980 at Chromosome 1A



Mapped EST Accession: **BE399980** [\[GrainGenes\]](#) | [\[NCBI\]](#) | [\[wEST-SQL\]](#) **Verified**

Contig Name: [NSFT03P2](#) [Contig18831](#)

Bin: [1AS1-0.47-0.86](#)

Forward Primer Name: [BE399980A_F1](#)

Reverse Primer Name: [BE399980_cpR1](#)

Chromosome/Genome: 1A

Ref Plant: *T. aestivum*, China (Ta21, ABD)

```

          10      20      30      40      50
ATAGATAATGGTATATACTACATTTAAGGATAATAAAACAAATGACAAAAC
TAGAATTTTATCTATGCAGTATCTGAAAAACAAAATTGCATCTTCAGAAAAC
GGAAAGACAGGATTGAAAACAGTTAAGCATGGTGAGCAAAATAGAAAAGCAAA
CATGCAACAGAGTGGTACAATGGTAGGCAAAACAGCAAGGTCTTCATTGCA
AGAGGATATTGATGCTTACCCATATACAAGTAGGTTTCATAAGTACAATA
TAAGCAGTTCTAGATATCAGAAGGCATCAGCATTATGGAAGACCAGGATG
ACATTACAGAAAACAGGATAGACCTAACCTTAGCATCCTCATCCCTGACAG
AGTGAAGGCGGAAGCGCCCTTGGTGTGCATACAGAAAGCCTGTAGTTCTCA
CCAGTCTTGGGAATGGAAATGATGTCTAATAGACAAAATGAAGATATCG
GCAATCAAGTGACGTTAAACAATCTTTGCAGAATGTGAATTCACATAGGAT
ACAATTAATTCTGTTACAAGATTAGAGCATACCCATGAAACCAGCAGGGT
AGGTCTTGTGAGTGGGACCTTTCCATCAACCATGATATGACGCTGCATA
AGGATAGACTGCACCTCACGGTAGGTGAGAGCATATTTTCAGTCTGTTCTC
CAGGATGAGGATCAGAGGAAGCCCTCCCTGGCCT [685 bases]
```

Exon Ranges: 328-426;533-685

Intron Ranges: 1-327;427-532

Lab: USDA



KANSAS STATE UNIVERSITY

Mining the allohexaploid wheat genome for useful sequence polymorphisms (2009, BBSRC, K Edwards)



cerealsDB.uk.net

Investigating gene function in cereals

Home

[BLAST Wheat genomic sequence](#)

[Download Wheat genomic sequence](#)

[Search DArT data](#)

[Deletion mapped DArT markers](#)

[Wheat database](#)

[SNPs](#)

[images at wheatbp](#)

[Monogram network](#)

SNPs Overview

Information on genetic relationship and hybrid development obtained using DNA marker are Single Nucleotide Polymorphisms (SNPs). SNPs (Single Nucleotide Polymorphisms) are a new generation of marker for genotyping. SNPs (Single Nucleotide Polymorphisms) are either transitions (A<>T or G<>C) or transversions (A<>C or G<>T) or developing methods for the high throughput detection of SNPs using a Single Nucleotide Primer Extension (SNUPE) assay.

- [Search cereal SNP data](#)
- [Browse wheat SNPs](#)
- [Browse barley SNPs](#)
- [Browse maize SNPs](#)
- [Browse rice SNPs](#)
- [Browse Saccharum SNPs](#)
- [Browse sorghum SNPs](#)

1. Generate 5X coverage of Chinese Spring
2. 5X coverage of complexity reduced cultivar DNA
3. Sequencing cDNA from different cultivars
4. Discover SNPs
5. Genotyping and mapping

Supported by the [BBSRC](#) and based at the [University of Bristol](#)

Maintained by [Gary Barker](#) Last updated May 2009



Accession	Cultivar	SNP	Allele	Count
A2	Chinese Spring	(1)
H1	Chinese Spring	(1)
E1	J1 K1 Chinese Spring	(3)
B2	E2 G2 kitaKEI1354	(3)
I2	kitaKEI1354	(1)
B1	D2 kitaKEI1354	(1) unknown (1)
D1	I1 N1 S1 F2 M2 Chinese Spring	(4) Valuevskaya (1) recital (1)
O1	L2 Chinese Spring	(1) Sumai3 (1)
L1	Chinese Spring	(1)
M1	Q1 U1 H2 Chinese Spring	(3) Valuevskaya (1)
C1	C2 N2 P2 kitaKEI1354	(1) recital (1) unknown (2)
W1	X1 DT4BCS	(2)
Y1	Z1 J2 DT4BCS	(2) Fidel (1)
O2	recital	(1)



Matt Hayden, DPI, Victorian AgriBiosciences Center, Australia

1. Re-sequencing about 1000 deletion-bin mapped genes from 24 wheat lines for SNP discovery
2. Genome sequence survey of hypomethylated restriction libraries (created using *Pst*I) for Australian wheat varieties. Overall, 30,000 tentative SNPs are identified and currently being validated
3. cDNA sequencing from 3 Australian cultivars

DIGITAL project, INRA, 2009-2011 (Catherine Feuillet, INRA)

1. ISBP-derived markers
2. Development of Illumina OPAs

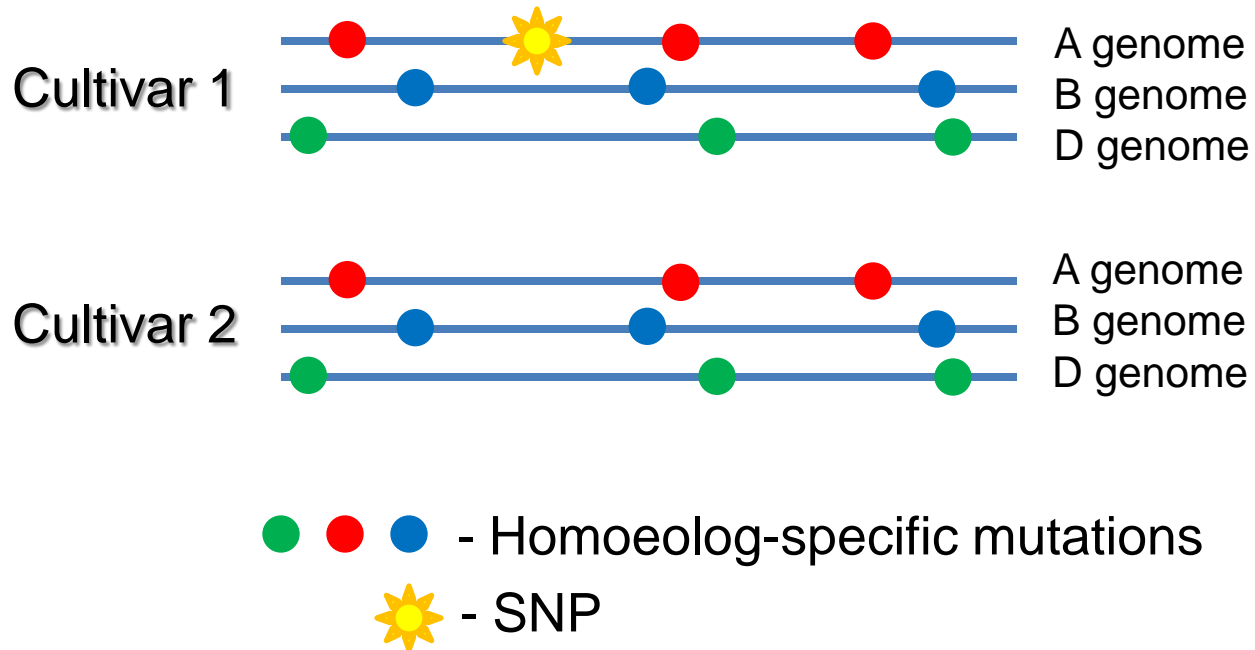
SNP markers for high-throughput genotyping to advance genomic, genetic and breeding research in wheat (USDA/AFRI Plant Genome, 2009-2012, E. Akhunov)

1. Perform a large-scale SNP discovery in a panel of 8 diverse US wheat lines using the next-generation sequencing instruments GS FLX (Roche) and Genome Analyzer (Illumina). Normalized cDNA libraries will be sequenced for the discovery of gene-associated SNPs in the wheat genome.
2. Develop Oligo Pool Assays (OPAs) for high-throughput genotyping with the GoldenGate assay (Illumina); validate ~ 9,200 SNPs by spring 2011.
3. Construct the SNP-based wheat genetic map and assess the patterns of linkage disequilibrium (LD) and genetic diversity in US wheat germplasm
4. Make the SNP resources and associated data developed in the project publicly accessible through the existing public databases GrainGenes and Gramene.

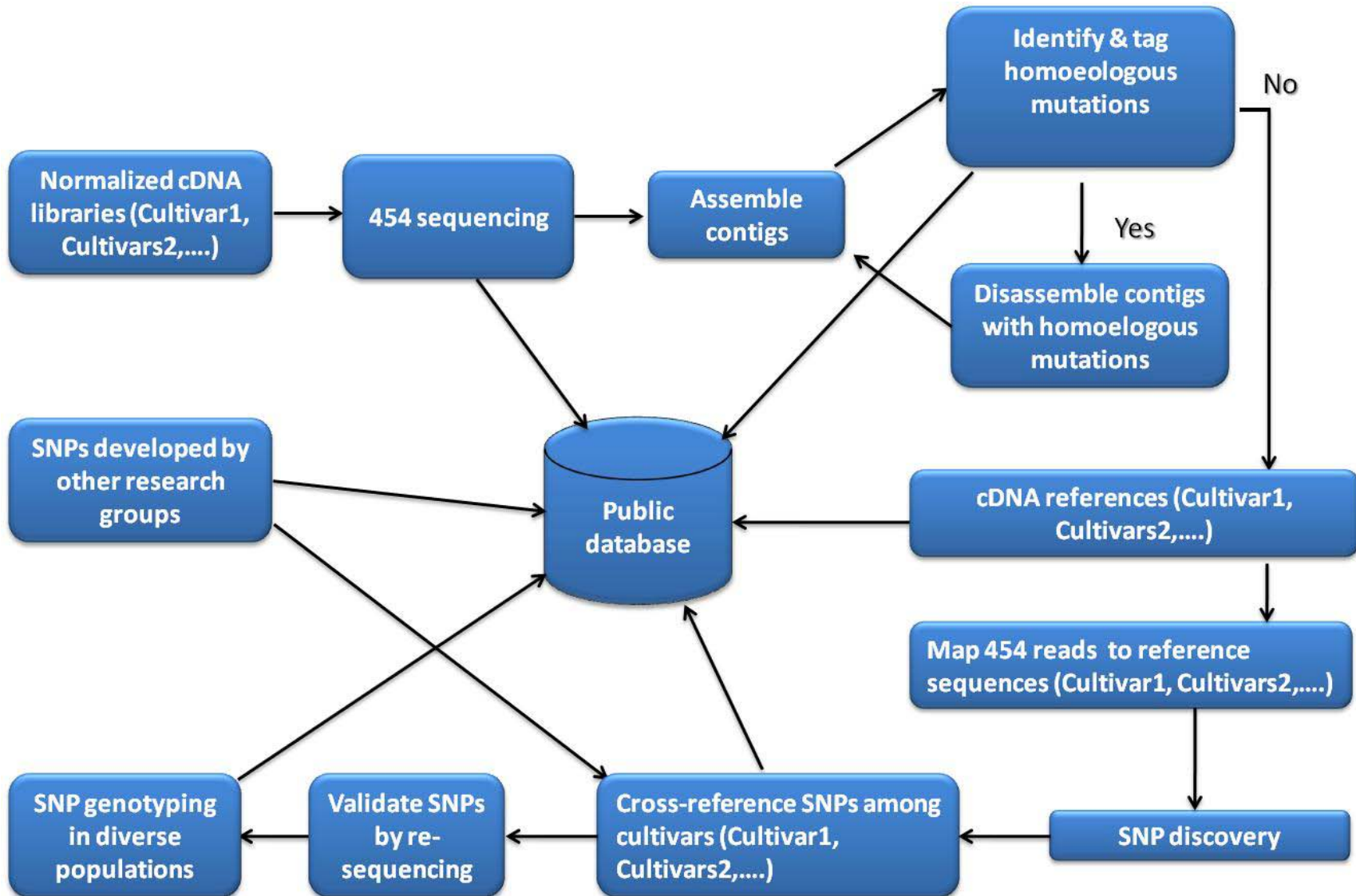
International wheat SNP working group

The primary goal of this working group is to facilitate the development of advanced open-access marker technologies based on SNPs by providing an organizational structure, communication and for making the results, both materials and information, available to broad wheat community.

SNPs in polyploid wheat



SNPs discovery workflow

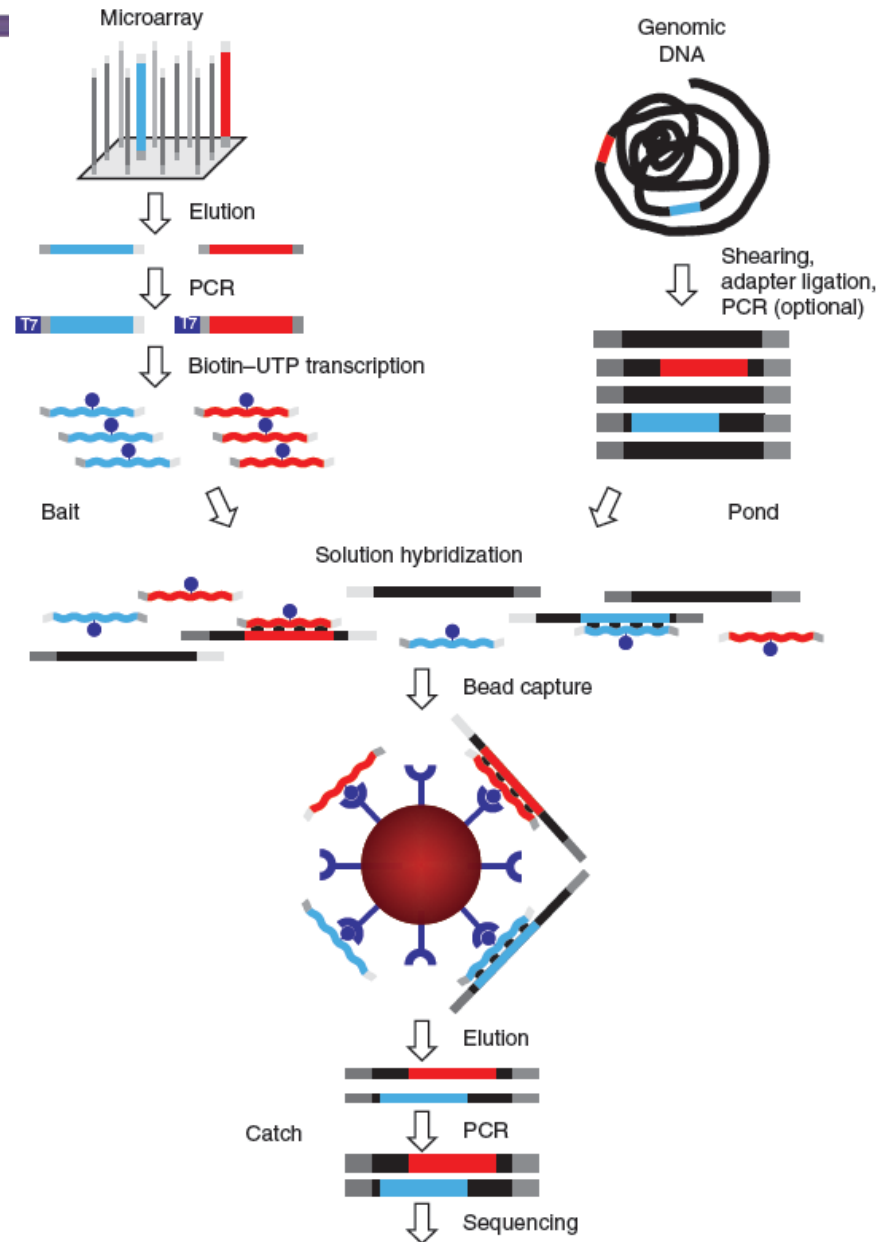


SNPs in polyploid wheat: discovery

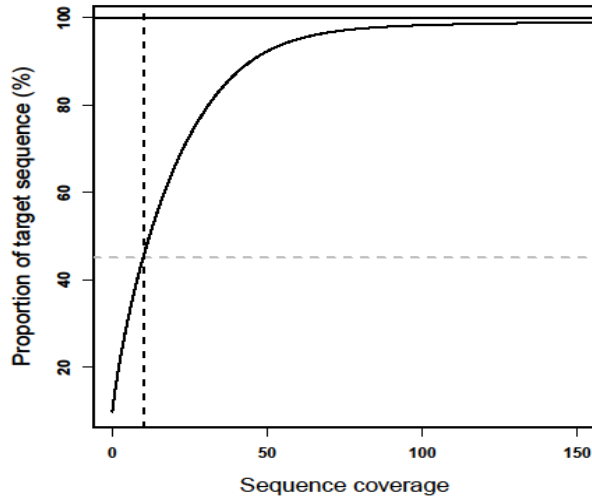
1. 454 and Illumina sequencing: 6 US cultivars + 3 Australian (~14 million reads)
2. Homoeolog-specific assembly (454 data): 120,000 contigs ~100 Mb
3. Map sequence reads to the reference: ~40% of reads can be mapped
4. SNP discovery: ~ 50,000 SNPs
5. Validation using Illumina OPA is underway
6. ~9,200 SNPs by spring 2011 will be tested by Illumina OPA genotyping

SNPs in polyploid wheat: sequence capture

The approach is tested on tetraploid wheat by sequencing DNA fragments captured from ~3,500 genes, 1 kb each



SNPs in polyploid wheat: sequence capture

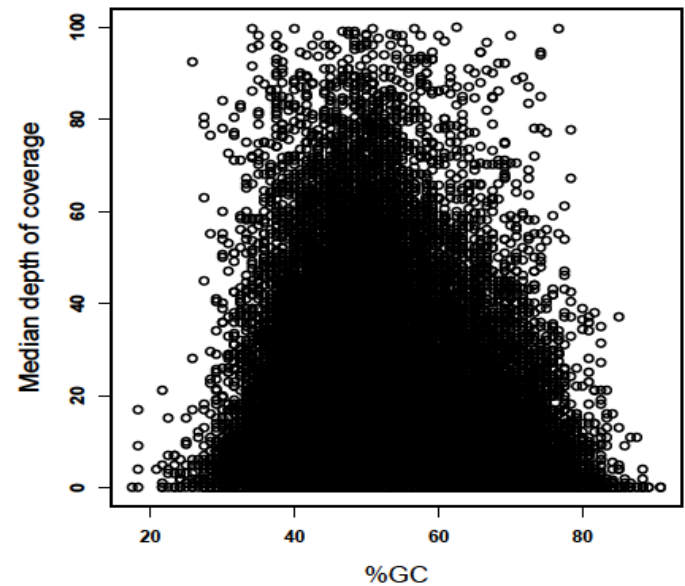


Cumulative distribution of coverage depth

A total of 42% of Illumina sequence reads were uniquely mapped;
A total of 90% of targeted sequences were covered by at least one read.

Relationship between median of coverage depth and capture bait GC-content

A total of **3,020** SNPs differentiating wild and domesticated emmer; 1 SNP every 1 kb of coding sequence.



Genome wide SNPs for polyploid wheat

- High resolution genetic maps
- Gene mapping using association mapping and nested associated mapping panels
 - *Use GW sets of SNPs for gene mapping*
 - *Markers showing associations can be used in MAS*
- Genomic selection
 - *Use genome wide SNPs for predicting phenotypes*
 - *Once incorporated in breeding scheme can speed up the development of new cultivars*

SNP genotyping in polyploid wheat: OPA

Activate DNA sample
(minimum 250 ng at 50 ng/uL)

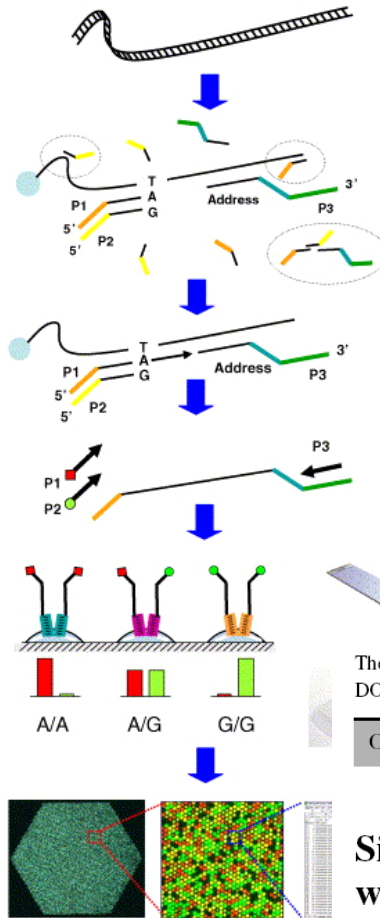
Hybridize assay
oligonucleotides and bind DNA
sample to paramagnetic
particles (mishybridized
oligonucleotides are circled)

Wash away excess and
mishybridized assay
oligonucleotides. Perform allele-
specific extension, then ligation
reaction.

PCR amplify extended and
ligated products with
fluorescently labeled primers.
Make PCR product single-
stranded.

Hybridize to Sentrix Array

Image Sentrix Array, generate
genotypes



Illumina Allele Specific
Primer Extension
(ASPE) and Ligation

Cy3- and Cy5-labeled
universal primers

Theor Appl Genet (2009) 119:507–517
DOI 10.1007/s00122-009-1059-5

ORIGINAL PAPER

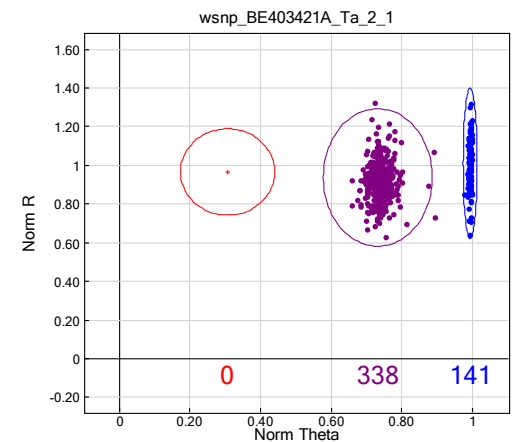
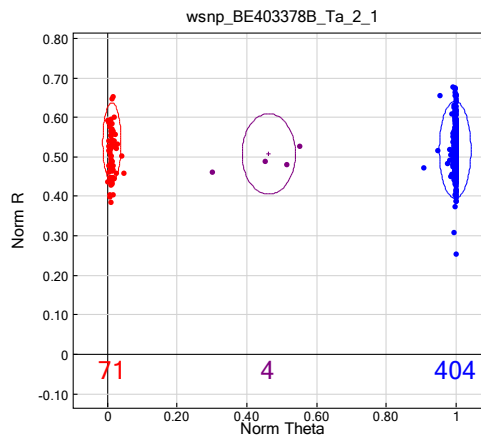
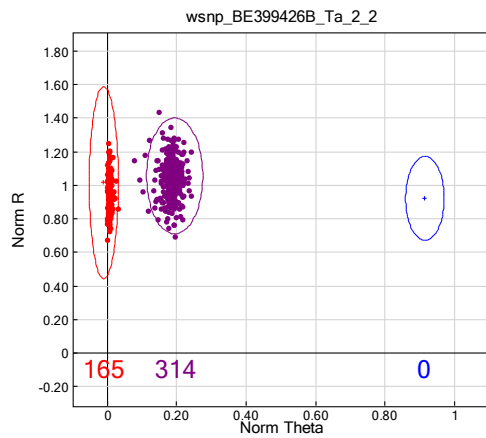
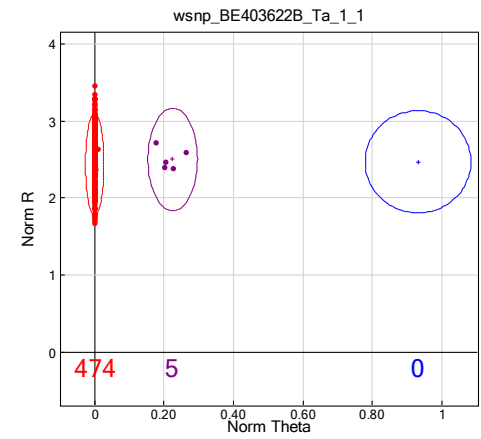
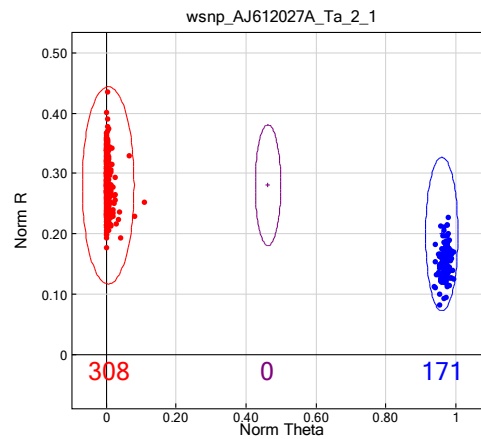
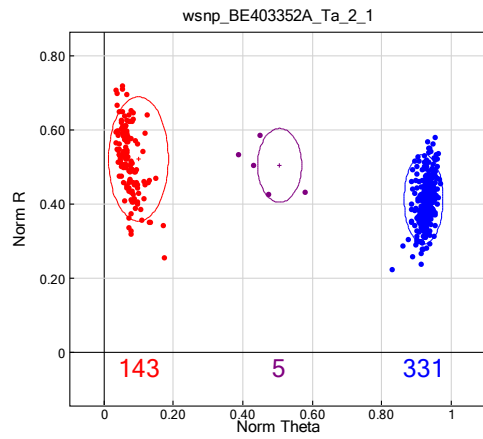
Single nucleotide polymorphism genotyping in polyploid wheat
with the Illumina GoldenGate assay

Eduard Akhunov · Charles Nicolet · Jan Dvorak

Wheat genotyping with 1536-plex wheat OPA

240 – spring wheat cultivars

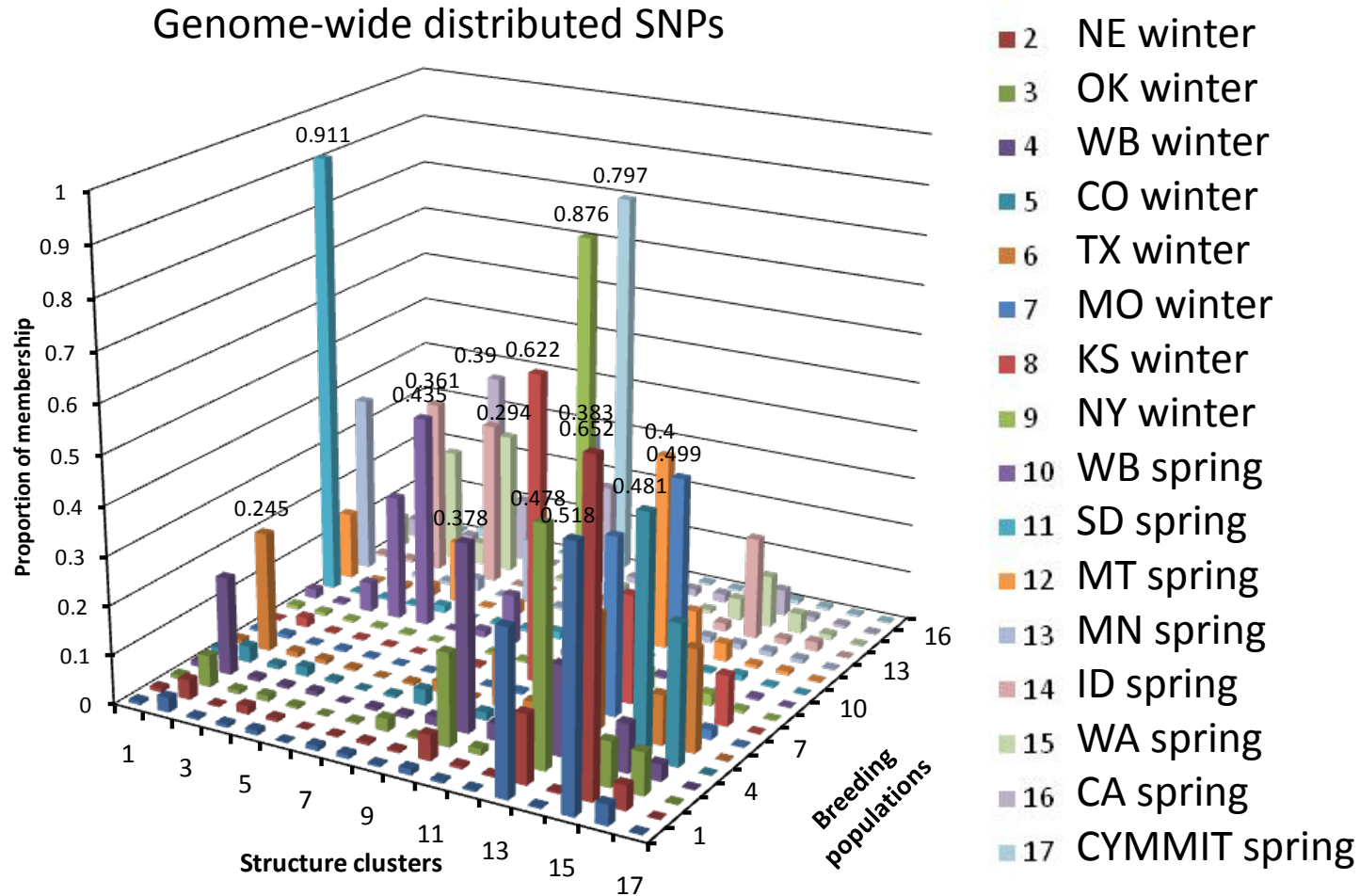
239 – winter wheat cultivars



SNP diversity within breeding populations

Growth Habit	Origin	Sample size	Polymorphic SNPs, (%)
winter	SD	21	79.5
	NE	49	84.8
	OK	40	84.8
	WestBread	11	78.4
	CO	30	82.7
	TX	38	85
	MT	34	82.6
	KS	4	64.9
	NY	10	70.3
		237	89.2
spring	WestBread	30	82.4
	SD	40	80.7
	MT	30	85
	MN	40	84.8
	ID	30	83.2
	WA	10	78.1
	CA	30	87
	CYMMIT	31	85.2
		241	89.1

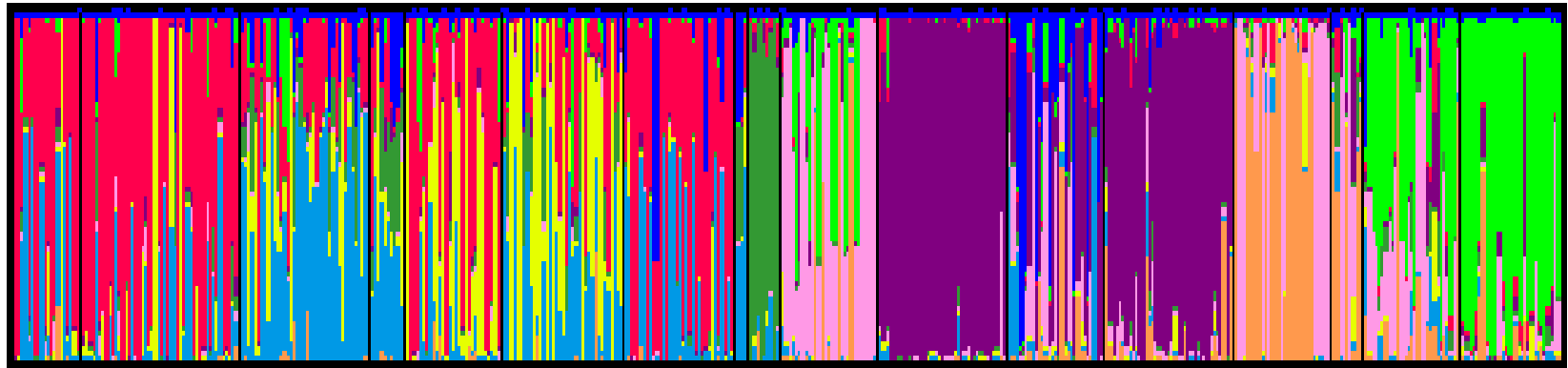
Proportion of membership of 17 breeding populations in 17 clusters inferred by STRUCTURE



Analysis of clustering suggest that the total number of subpopulations in our sample should not exceed 9 (4 subpopulations in winter wheat panel; 5 subpopulations in spring wheat panel).

Population structure of US wheat

Number of clusters K=9



SD winter

NE winter

OK winter

WB winter

CO winter

TX winter

MO winter

KS winter

NY winter

WB spring

SD spring

MT spring

MN spring

ID spring

WA spring

CA spring

CY spring

Community resource: the panel of winter wheat lines is being increased now and will be soon available for association mapping experiments

Genome-wide association mapping (GWAM)

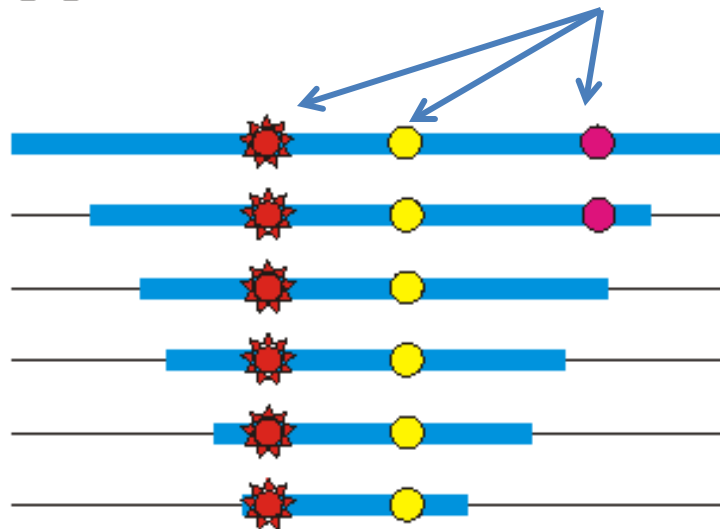
GWAM – search for statistically significant association between a query SNP and a causal polymorphism, or is in strong LD with the causal site.

Query SNP = Causal polymorphism

Query SNP is in LD with causal polymorphism

☀ Causal mutation
●● SNPs

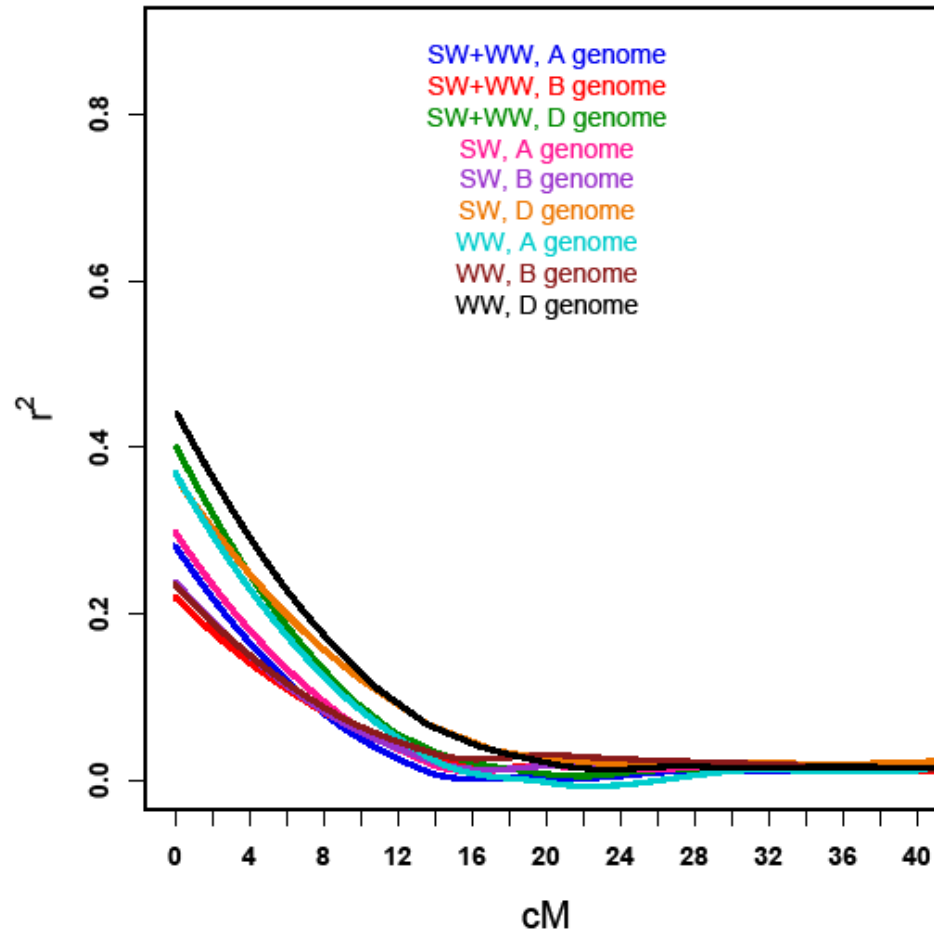
Complete LD at the time of origin



Factors decreasing LD:
Recombination

Factors increasing LD:
Drift
Selection
Admixture

Rate of linkage disequilibrium (LD) decay



Association mapping using winter wheat panel

Yellow rust, seedling stage resistance

237 lines, FDR > 0.3

Locus	P_Value	FDR
SNP248	0.0010	0.11
SNP284	0.0006	0.10
SNP353	0.0001	0.08
SNP456	0.0006	0.10
SNP599	0.0016	0.16
SNP822	0.0004	0.10

139 lines, Q > 0.5, FDR > 0.3

Locus	P_Value	FDR
SNP92	0.0023	0.26
SNP248	0.0007	0.23
SNP284	0.0022	0.26
SNP353	0.0015	0.25
SNP456	0.0022	0.26
SNP822	0.0015	0.25

96 lines, Q > 0.7, FDR > 0.1

Locus	P_Value	FDR
SNP44	0.0009	0.06
SNP92	0.0014	0.08
SNP212	0.0017	0.09
SNP307	3.1×10^{-23}	4.0×10^{-21}
SNP353	9.3×10^{-08}	1.0×10^{-05}
SNP360	0.0009	0.06
SNP381	1.2×10^{-24}	2.7×10^{-22}
SNP403	5.9×10^{-05}	0.01
SNP671	1.2×10^{-24}	2.7×10^{-22}
SNP751	3.1×10^{-23}	4.0×10^{-21}
SNP762	0.0005	0.04
SNP790	1.2×10^{-24}	2.7×10^{-22}

Association mapping in stratified samples

The extent of LD can be controlled by selecting panels of more closely related cultivars (for example, using membership coefficient Q)

Increase in LD offers the possibility of GWAM mapping using smaller set of SNP markers

AM mapping showed the genetic heterogeneity of disease resistance trait

Diverse panels offer higher resolution but require genotyping with more densely spaced SNP markers

Future of genotyping: Restriction site associated DNA / Genotyping by sequencing

OPEN ACCESS Freely available online

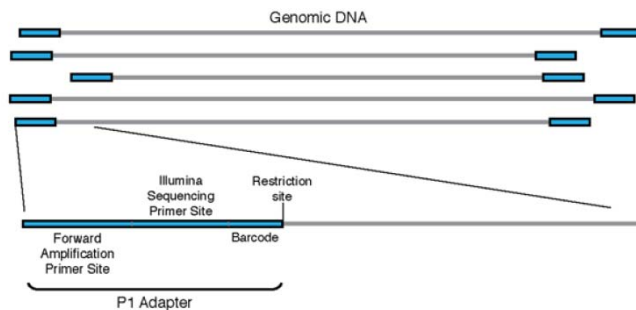
PLoS one

Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers

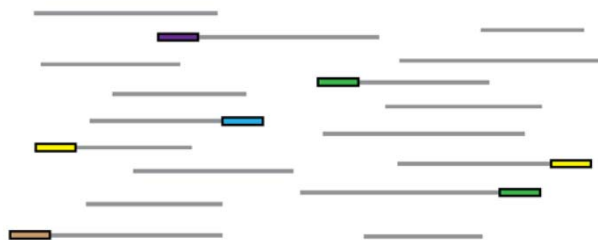
Nathan A. Baird¹*, Paul D. Etter¹*, Tressa S. Atwood², Mark C. Currey³, Anthony L. Shiver¹, Zachary A. Lewis¹, Eric U. Selker¹, William A. Cresko³, Eric A. Johnson¹*

¹ Institute of Molecular Biology, University of Oregon, Eugene, Oregon, United States of America, ² Floragenex, Eugene, Oregon, United States of America, ³ The Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, Oregon, United States of America

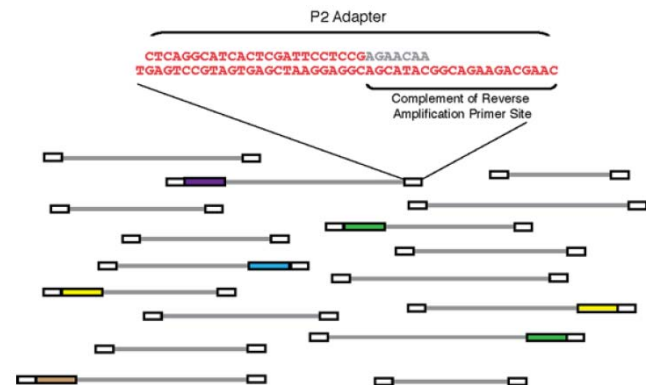
A Ligate P1 Adapter to digested genomic DNA



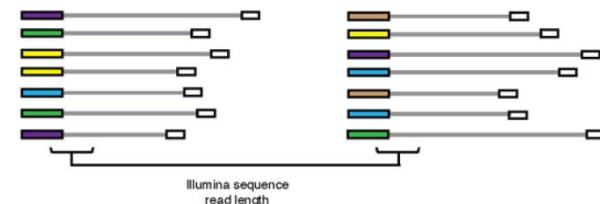
B Pool barcoded samples and shear



C Ligate P2 Adapter to sheared fragments



D Selectively amplify RAD tags



High-throughput sequencing systems: HiSeq2000



Workflow & Specs

HiSeq 2000 Preliminary Performance Parameters*

Read Length	Run Time	Output
1 x 35 bp	~1.5 days	26–35 Gb
2 x 50 bp	~4 days	75–100 Gb
2 x 100 bp	~8 days	150–200 Gb

*Sequencing output generated with a PhIX library and cluster densities between 260,000–347,000 clusters/mm² that pass filtering on a HiSeq 2000.

Throughput

Up to 25 Gb per day for a 2 x 100 bp run.

Reads

Up to one billion clusters passing filter, and up to two billion paired-end reads.

Service and Support

Illumina will ensure that your HiSeq 2000 is properly installed and qualified, and will provide ongoing maintenance and service. This industry-leading support is available in North America, Europe, and Asia.

Make possible development of high-density sequence based genetic maps by sequencing mapping populations;

Offers possibility of performing GWAM in diverse panels of lines

Conclusions:

Extensive resources are being developed for high-throughput analysis of genetic variation in wheat. Resource development efforts need to be coordinated to increase their value.

GWAM in wheat is close to become reality.

Even currently available sets of SNP markers can be used for GWAM by assembling panels from genetically less diverse cultivars (regional panels). These sets of markers can already be used for training prediction models for genomic selection.

New approaches and technologies allow cost-effective analysis of genetic diversity at ever increasing level of precision with the possibility in near future to study diversity at the whole genome sequence level. Accumulation of high-quality phenotypic data will be critical for understanding the genetic basis of many important traits in wheat.



Acknowledgements

C. Saintenac (KSU, Manhattan)
V. Catana (KSU, Manhattan)
A. Akhunova (KSU, Manhattan , IGF)
R. Matnyazov (KSU, Manhattan)
S. Kiani (KSU. Manhattan)

J. Dvorak (UC, Davis)
M.C. Luo (UC, Davis)
J. Dubcovsky (UC, Davis)
P.S. Baenziger (UN, Lincoln)

S. Haley (CSU, Fort Collins)
D.R. Clark (WestBread, Bozeman)
L. Talbert (MSU, Bozeman)
J.A. Anderson (UMN, St. Paul)

S. Dreisigacker (CYMMIT)
K. Glover (SDSU, Brookings)
J. Chen (UI, Aberdeen)
K. Campbell (USDA, Pullman)
P.L .Bruckner (MSU, Bozeman)
J.C. Rudd (TAMU, Amarillo)
B.F. Carver (OSU, Stillwater)
R. Sims (WestBread, Haven)
M.E. Sorrells (Cornell, Ithaca)

S. Chao (USDA ARS, Fargo)

