

Genotyping – By – Sequencing (GBS) in Crop Plants

Trushar Shah &
Rajneesh Paliwal



RESEARCH
PROGRAM ON
DrylandCereals

Topics presented

GBS OVERVIEW AND SNP CALLING

GENOME WIDE ASSOCIATION STUDY (GWAS)

APPLICATION OF GBS-SNP INCLUDING GWAS

Genotyping by sequencing (GBS) in any large genome species requires reduction of genome complexity

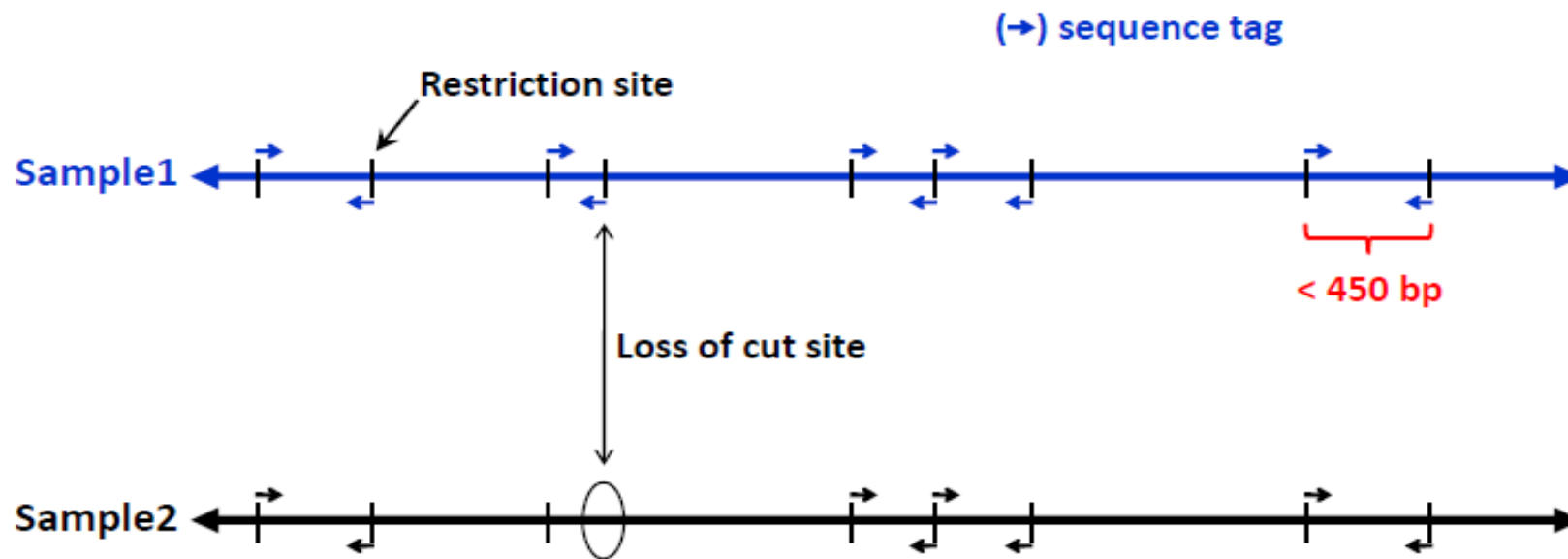
Target enrichment

- Long range PCR of specific genes or genomic subsets
- Molecular inversion probes
- Sequence capture approaches hybridization-based (microarrays)

Restriction Enzymes (REs)

- ***Technically less challenging***
- Methylation sensitive Res filter out repetitive genomic fraction

Overview of Genotyping by Sequencing (GBS)



- Focuses NextGen sequencing power to ends of restriction fragments in both reference and non-reference genome plants
- Both SNPs and presence/absence markers can be scored
- Small Indels are identified but are not scored

GBS is a simple, highly multiplexed system for constructing libraries for NGS

- Reduced Sample handling**
- Few PCR & purification steps**
- No DNA size fractionation**
- Efficient barcoding system**
- Simultaneous marker discovery and genotyping**
- Scales very well**

Most frequently asked question for new species

“How many SNP will I get?”

Answer: “It depends.....”

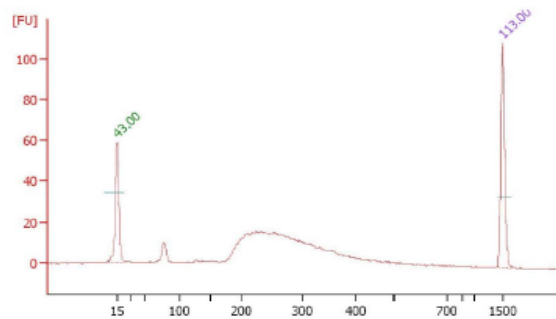
- Genome size and expected heterozygosity affects size of fragment pool for desired amount of sequence coverage (enzyme choice and multiplex level).
- Amount of extent diversity and how well your samples capture that diversity
- Reference genome sequence? 3-4x more SNPs attained by aligning to a reference sequence

Cont...

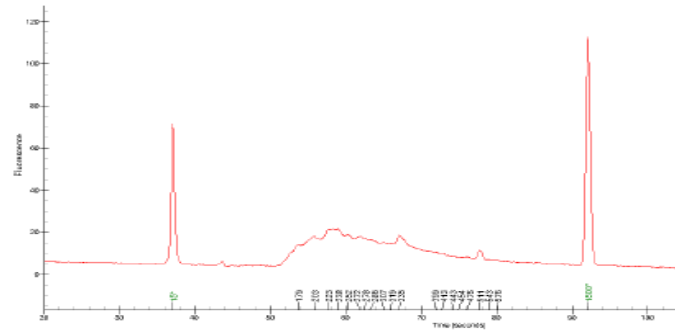
Species	Genome Size (Mb)	Enzyme	Sample Size	No. SNPs
Maize	2,600	<i>ApeKI</i>	33,000	2,200K
Rice	400	<i>ApeKI</i>	850	60K
Grape	500	<i>ApeKI</i>	8,000	1,200K
Willow*	460	<i>ApeKI</i>	459	23K
Pine*	16,000	<i>ApeKI</i>	12	63K
Vole*	3,400	<i>PstI</i>	283	53K
Fox*	2,400	<i>EcoT22I</i>	48	16K
Cow	3,000	<i>PstI</i>	48	64K
<i>Neurospora</i> (fungus isolates)	40	<i>ApeKI</i>	384	100K

*No reference genome. UNEAK analysis pipeline used for analysis. To avoid homology/paralogy issues this pipeline calls SNPs very conservatively.

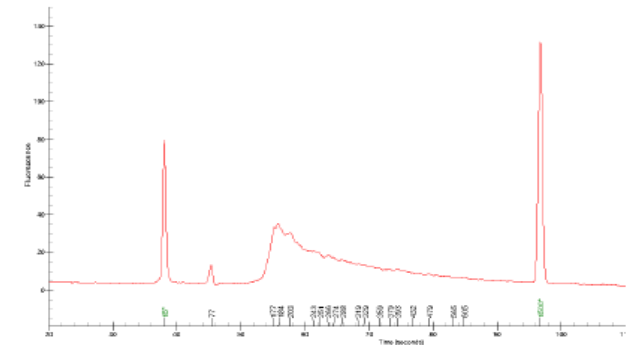
PstI works well for mammals and birds.



EcoT22I works well for fish, amphibians and invertebrates.



ApeKI works well for grasses.



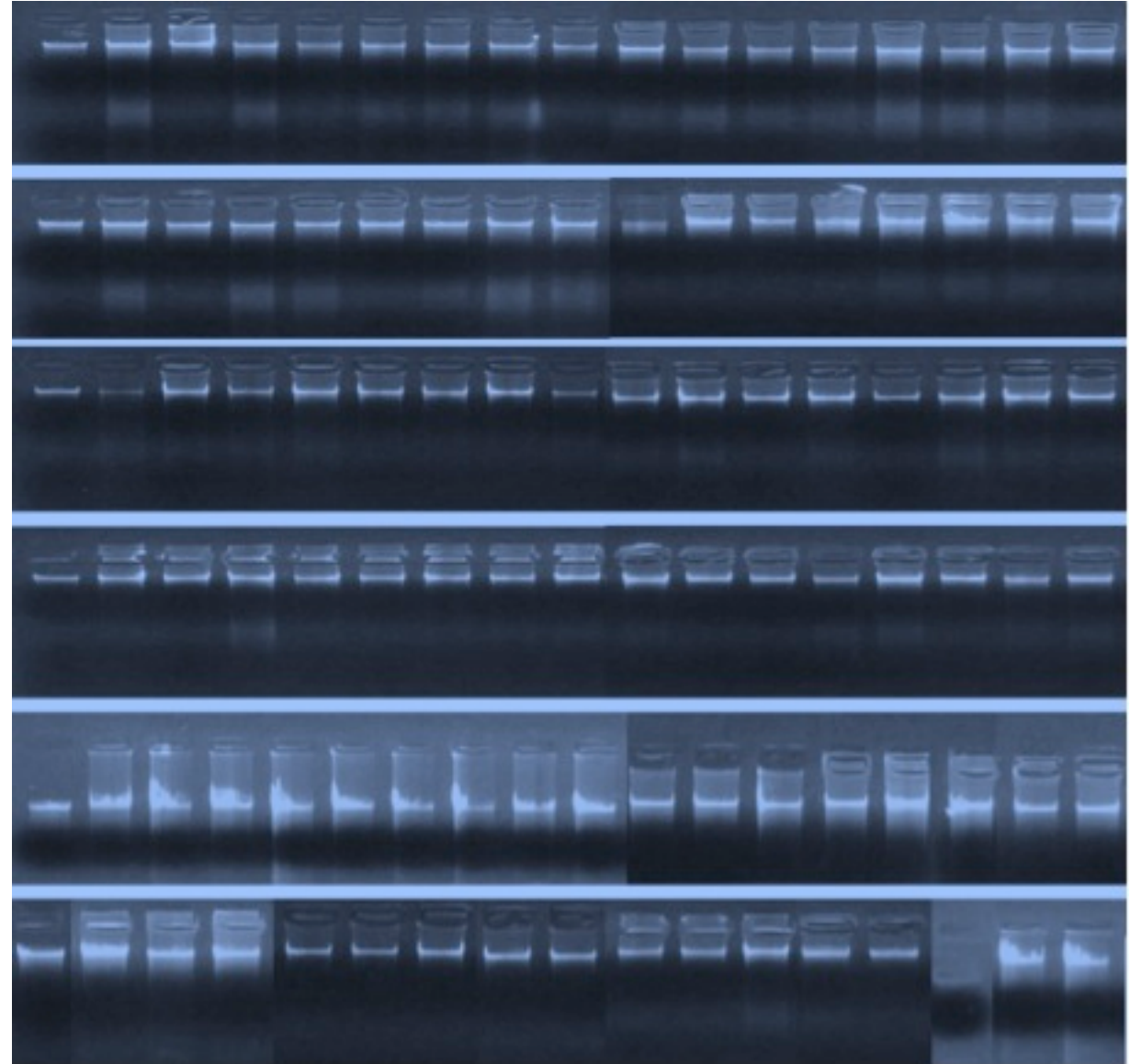
Most significant GBS technical issues?

- DNA Quality
- DNA quantification

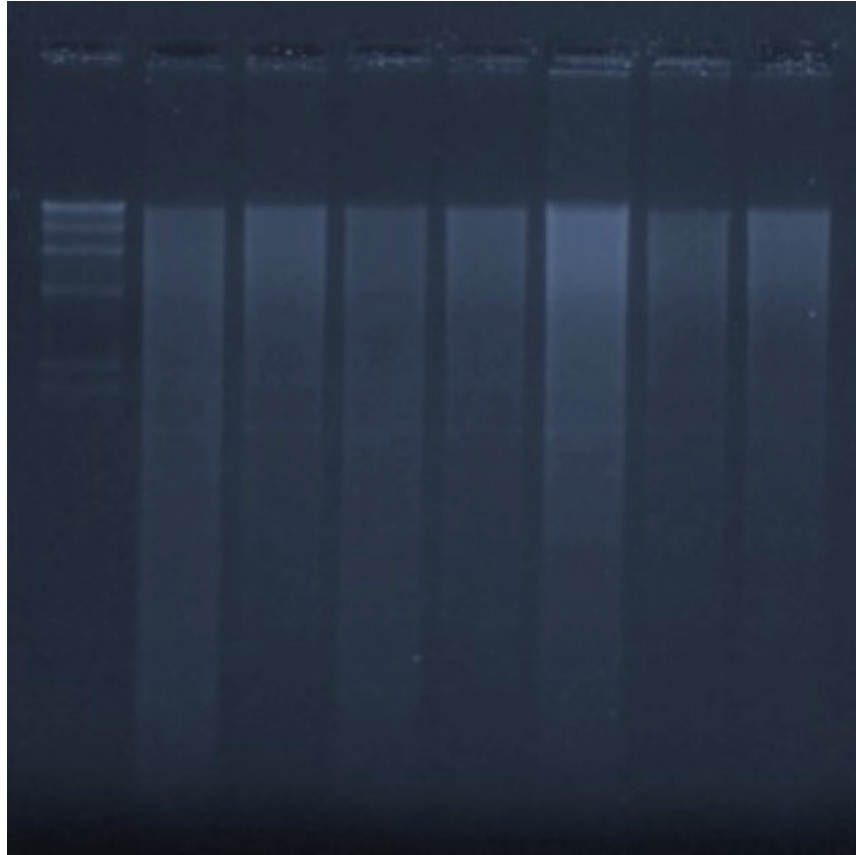
SAMPLE PROCESSING

DNA – Next Generation Sequencing

- High molecular weight
- Good quality

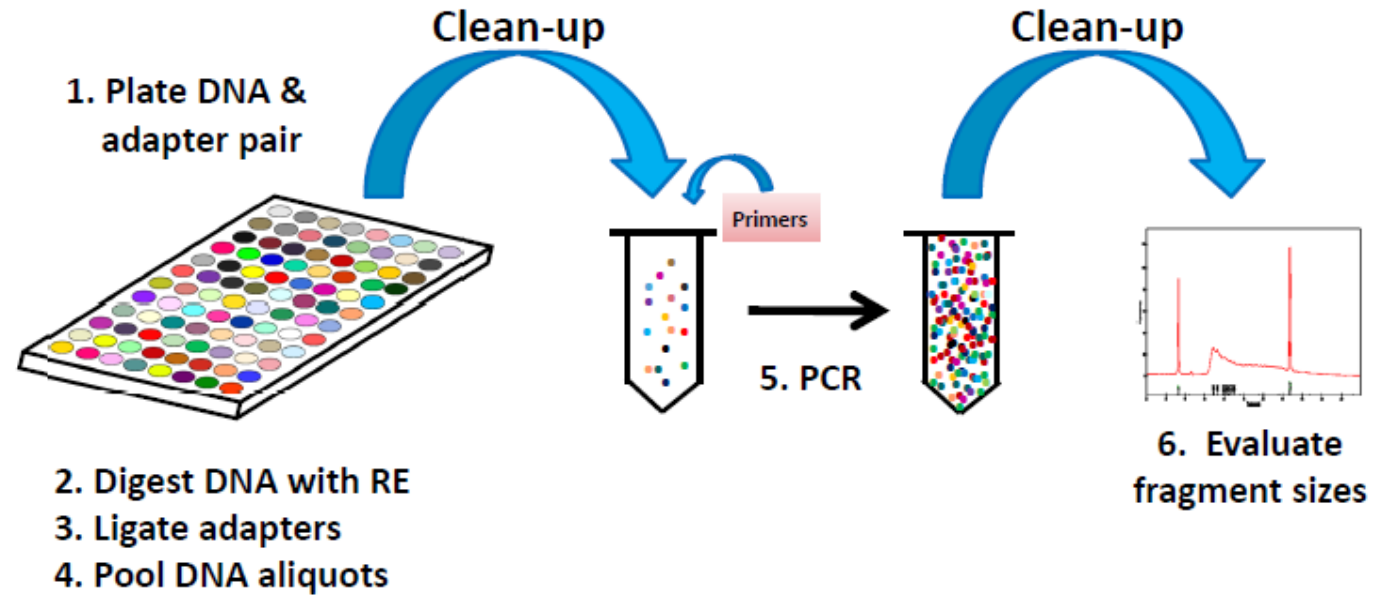


Restriction Digestion

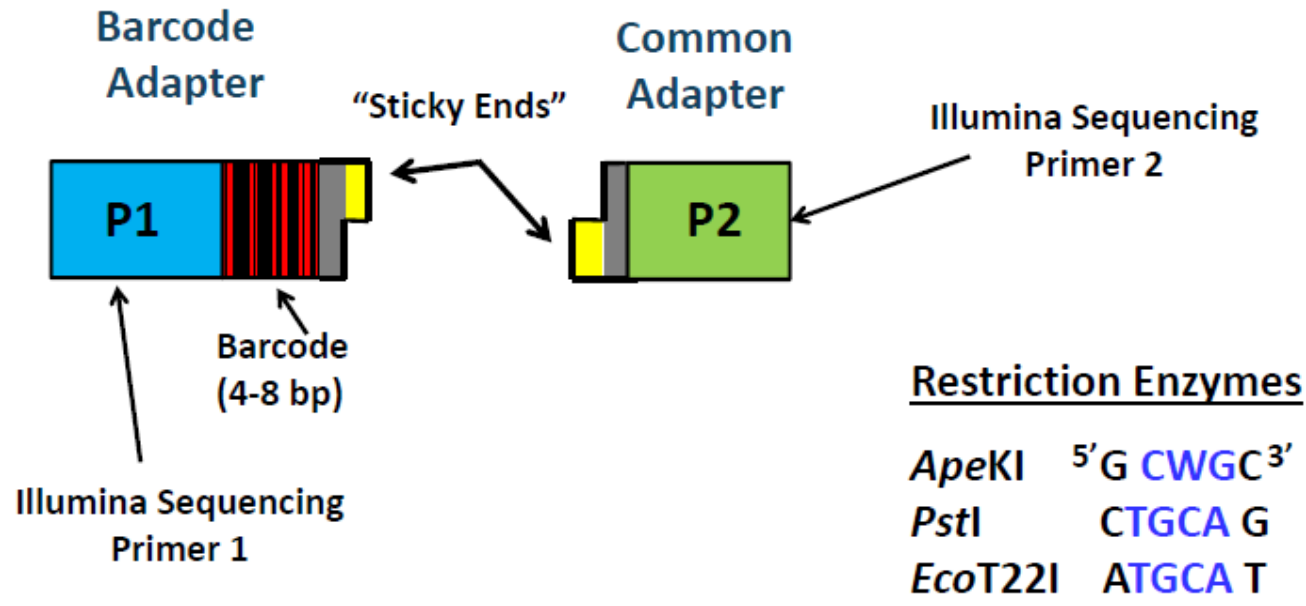


GBS 96- or 384-plex Protocol

(<http://www.maizegenetics.net/gbs-overview>)

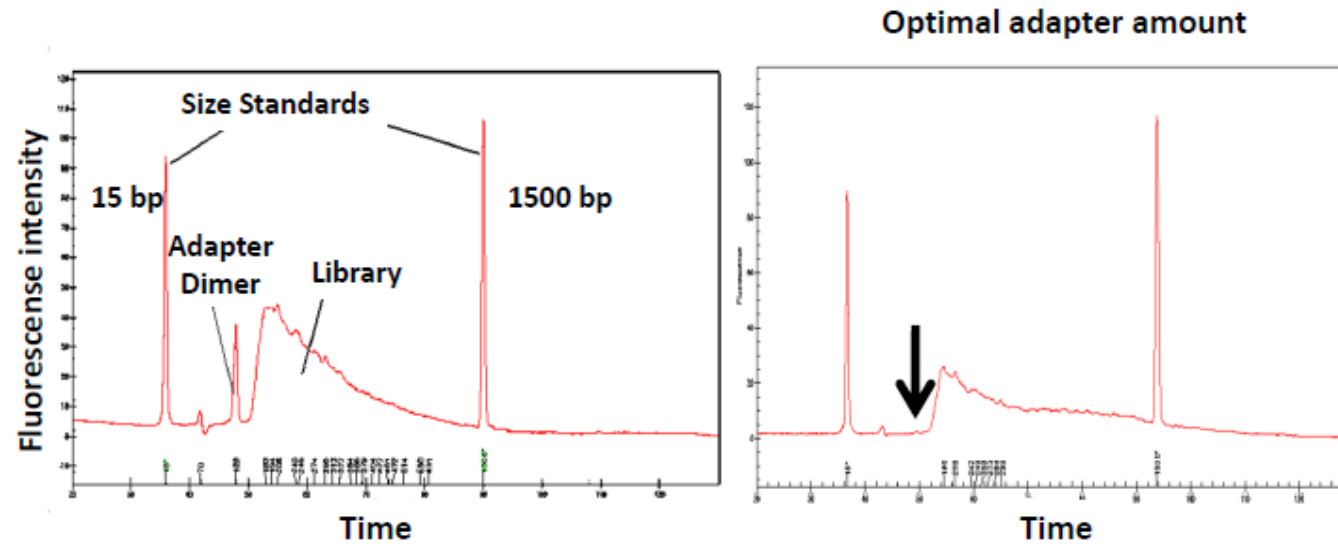


GBS Adapters and Enzymes

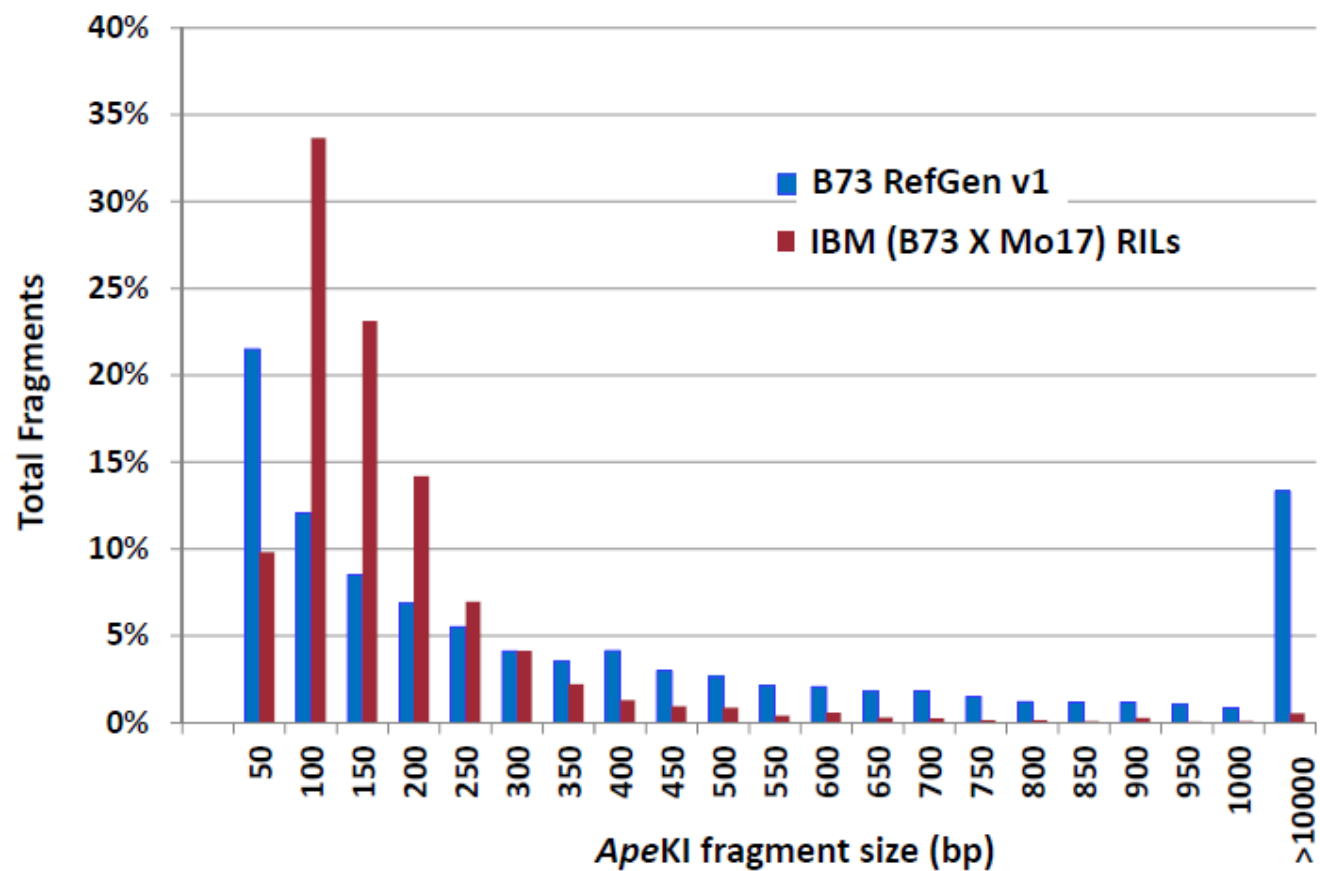


Perform Titration to Minimize Adapter Dimers Before Sequencing

NOTE: Done once with a small number of samples.
Adapter dimers constitute only 0.05% of raw sequence reads

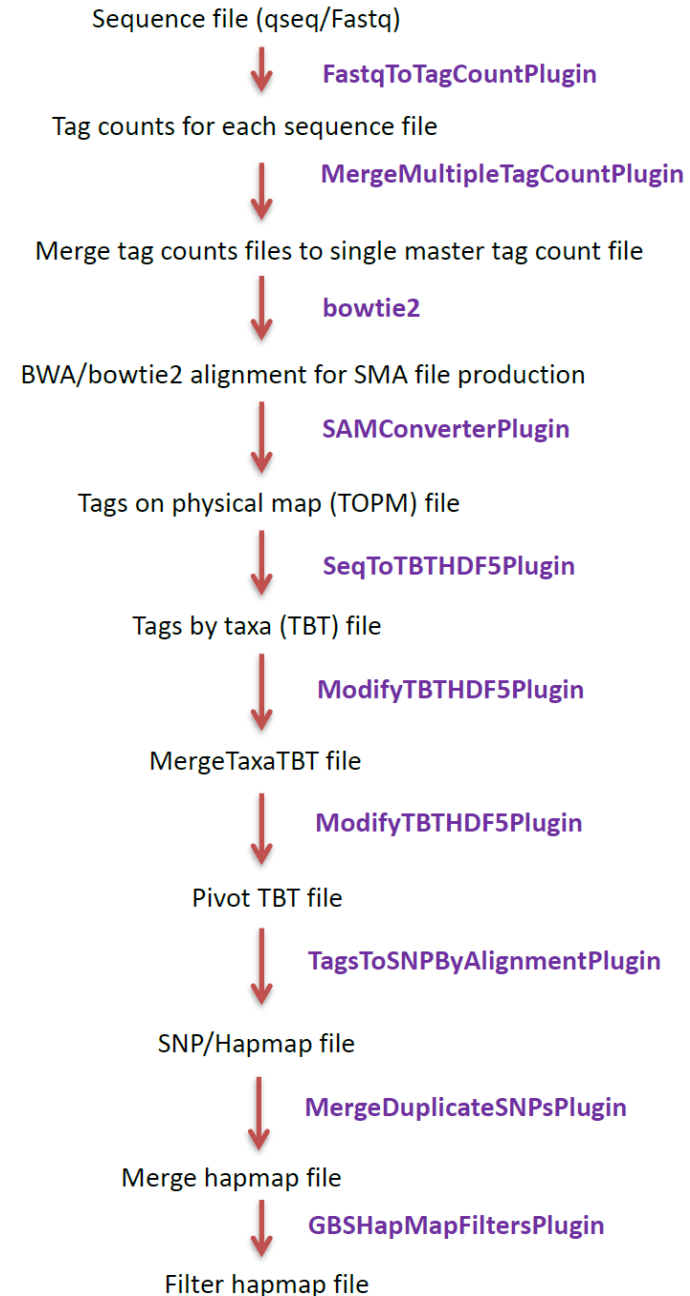


Small Fragments are Enriched in GBS Libraries



GBS Work flow:

Bioinformatics (GBS Pipelines)



GBS pipelines

TASSEL – tassel4.0standalone

http://www.maizegenetics.net/index.php?option=com_content&task=view&id=89&Itemid=119

User manual: Reference genome

<http://www.maizegenetics.net/tassel/docs/TasselPipelineGBS.pdf> (Elshire et al. 2011)

User manual: without reference genome

http://www.maizegenetics.net/images/stories/bioinformatics/TASSEL/uneak_pipeline_documentation.pdf (Lu et al. 2013)

For more information on GBS

<http://www.maizegenetics.net/gbs-bioinformatics>

Computationally intensive

JAVA environment

PERL + JAVA

GBS Vocabulary

Taxa	Individual sample
Key file	Text file containing <ul style="list-style-type: none">- Sample information- Barcode- Flowcell and lane number- Sample prepID
Barcode	Unique DNA sequence associated with each taxa
Sequence file	Text file containing DNA sequences information from Illumina <ul style="list-style-type: none">- Qseq or Fastq file
Read	DNA Sequence produced in sequencing..
GBS Tag	DNA sequence starts with cut site remnant and having additional sequence without
barcode	
TagsByTaxa (TBT)	Matrix of GBS tags (row) with taxa (columns)

GBS pipelines

Types

Discovery pipeline

- ❖ Requires a reference genome
- ❖ Multiple steps to process the data into genotypes

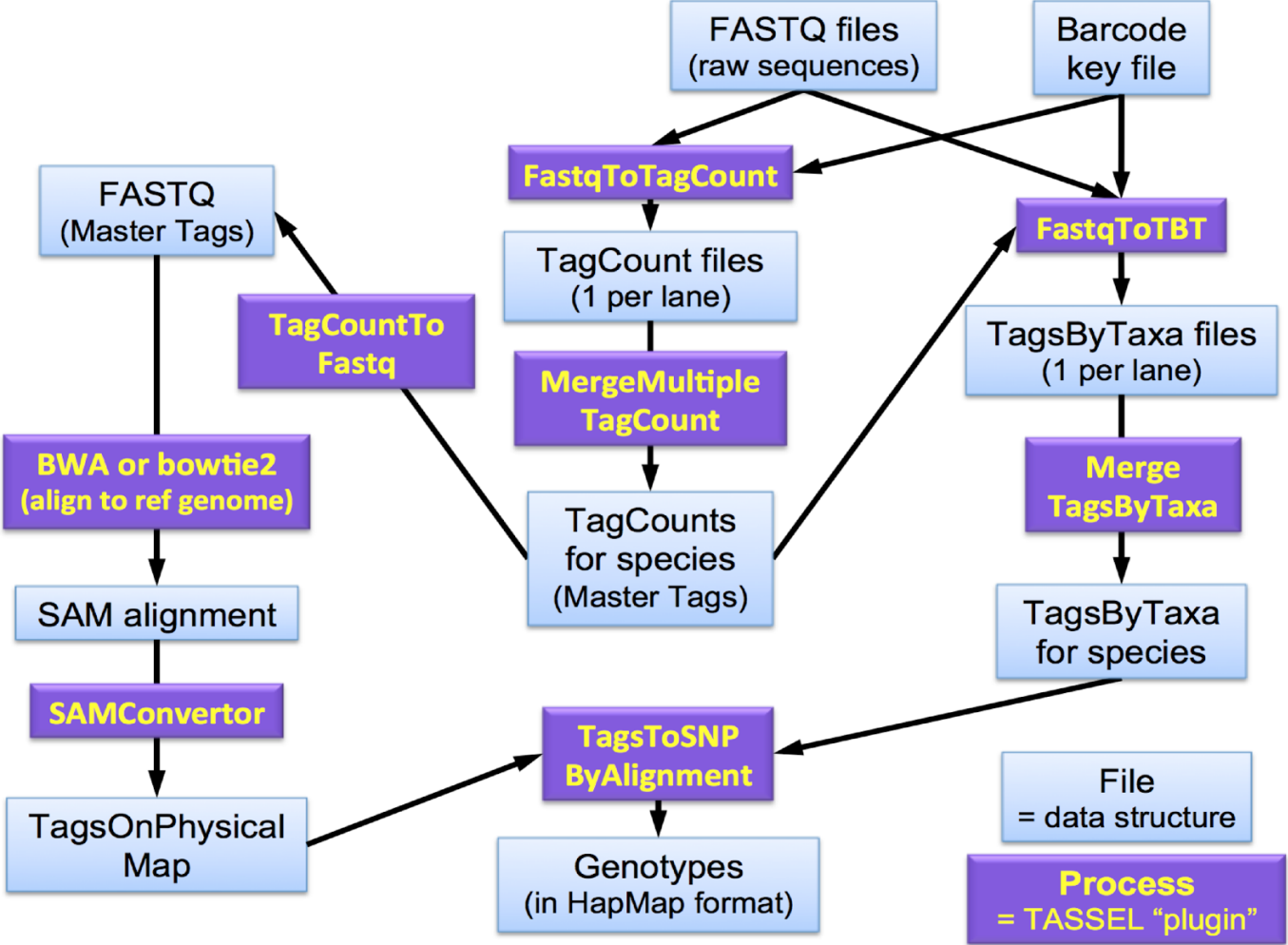
Production pipeline

- ❖ Uses info from discovery pipeline
- ❖ Only one step from sequence to genotypes

UNEAK pipeline

- ❖ Species without reference genome

Discovery GBS Pipeline



Discovery pipeline

Key file format

Flowcell	Lane	Barcode	DNA Sample	LibraryPlate	Row	Col	Library PrepID	Library PlateID	Enzyme	FullSampleName
628NVAAXX	5	TGCA	IS2787	ICRISAT Plate 1A	A	2	250006891	450013296	ApeKI	IS2787:628NVAAXX:5:250006891
628NVAAXX	5	ACTA	IS3957	ICRISAT Plate 1A	A	3	250006893	450013296	ApeKI	IS3957:628NVAAXX:5:250006893
628NVAAXX	5	GTCT	IS6193	ICRISAT Plate 1A	A	4	250006895	450013296	ApeKI	IS6193:628NVAAXX:5:250006895
628NVAAXX	5	GAAT	IS9303	ICRISAT Plate 1A	A	5	250006897	450013296	ApeKI	IS9303:628NVAAXX:5:250006897
628NVAAXX	5	GCGT	IS11119	ICRISAT Plate 1A	A	6	250006899	450013296	ApeKI	IS11119:628NVAAXX:5:250006899
628NVAAXX	5	TGGC	IS13845	ICRISAT Plate 1A	A	7	250006901	450013296	ApeKI	IS13845:628NVAAXX:5:250006901
628NVAAXX	5	CGAT	IS15752	ICRISAT Plate 1A	A	8	250006903	450013296	ApeKI	IS15752:628NVAAXX:5:250006903
628NVAAXX	5	CTTGA	IS19132	ICRISAT Plate 1A	A	9	250006905	450013296	ApeKI	IS19132:628NVAAXX:5:250006905
628NVAAXX	5	TCACC	IS20351	ICRISAT Plate 1A	A	10	250006907	450013296	ApeKI	IS20351:628NVAAXX:5:250006907
628NVAAXX	5	CTAGC	IS22330	ICRISAT Plate 1A	A	11	250006909	450013296	ApeKI	IS22330:628NVAAXX:5:250006909
628NVAAXX	5	ACAAA	IS23669	ICRISAT Plate 1A	A	12	250006911	450013296	ApeKI	IS23669:628NVAAXX:5:250006911
628NVAAXX	5	TTCTC	IS303	ICRISAT Plate 1A	B	1	250006913	450013296	ApeKI	IS303:628NVAAXX:5:250006913
628NVAAXX	5	AGCCC	IS2807	ICRISAT Plate 1A	B	2	250006915	450013296	ApeKI	IS2807:628NVAAXX:5:250006915
628NVAAXX	5	GTATT	IS3971	ICRISAT Plate 1A	B	3	250006917	450013296	ApeKI	IS3971:628NVAAXX:5:250006917
628NVAAXX	5	CTGTA	IS6351	ICRISAT Plate 1A	B	4	250006919	450013296	ApeKI	IS6351:628NVAAXX:5:250006919
628NVAAXX	5	AGCAT	IS9468	ICRISAT Plate 1A	B	5	250006921	450013296	ApeKI	IS9468:628NVAAXX:5:250006921
628NVAAXX	5	ACTAT	IS12169	ICRISAT Plate 1A	B	6	250006923	450013296	ApeKI	IS12169:628NVAAXX:5:250006923
628NVAAXX	5	GAGAAT	IS13926	ICRISAT Plate 1A	B	7	250006925	450013296	ApeKI	IS13926:628NVAAXX:5:250006925
628NVAAXX	5	CCAGCT	IS16044	ICRISAT Plate 1A	B	8	250006927	450013296	ApeKI	IS16044:628NVAAXX:5:250006927

GBS Restriction Fragment Structure



Accepted read



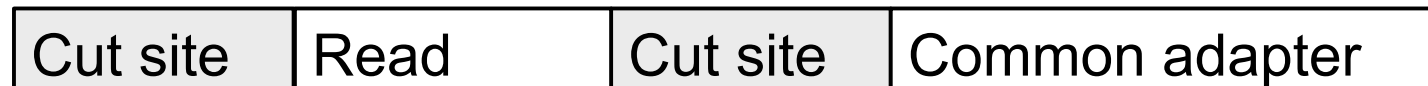
ATGCCTCTGCCTCGTTTAAGTCCGATTGCTGATAGTTATCTTTGTTTCTTACCAAATCCACATGGCTG

Rejected or Trimmed reads

Potential chimeric sequence



Short sequence



Adapter dimer



Raw sequence data

Read

ATGCCTCTGCCCTCGTTTAAGTCCGATTGCTGATAGTTATCTTTGTTTCTTACCAAATCCACATGGCTGAGATCGGAAGAGCGGT



GBS pipelines



ATGCCTCTGCCCTCGTTTAAGTCCGATTGCTGATAGTTATCTTTGTTTCTTACCAAATCCACATGGCTGAGATCGGAAGAGCGGT
CGATCAGCAGTTGACTGGACATCTAGGGGCAAAGCACTGTTTCGGTGCGGGCTGAGATCGGAAGAGCGGTTTCAGCAGGAATGCCGA
CCGGATATCAGCAGGCCGTGGTACATGTAATGGAGCATGGATTGAAGGTGGATGCCTTCATGTCTGGACGGCGATCGAGCTCGC



TagsCounts

CTGCCCTCGTTTAAGTCCGATTGCTGATAGTTATCTTTGTTTCTTACCAAATCCACATGGCTG
CAGCAGTTGACTGGACATCTAGGGGCAAAGCACTGTTTCGGTGCGGGCTGAAAAAAAAA
CAGCAGGCCGTGGTACATGTAATGGAGCATGGATTGAAGGTGGATGCCTGCAAAAAAAAA

Tag count > 5/10

MergeMultipleTagCountPlugin



Merge Tags counts

MasterTagCounts

Total # Tags	Tags size
1453006 2	
CTGCCCTCGTTTAAGTCCGATTGCTGATAG...GTTTCTTACCAAATCCACATGGCTG	64 409
CAGCAAAAAAAAAAAGGACATGGGTATCCGGTAACTGCAGAAAATTTGAGTAAAAGGCGGTA	64 314
CAGCAAAAAAAAAAATCTCAGAAAATGCAGTGCAGAGTGATATTTTGCTTGATCCTGGCTCAC	64 235
CAGCAAAAAAAAAACAAAAAACAAAAACAAAGGATTTCTTTTAGGGAAAAACATAAGGTGCG	64 338
CAGCAAAAAAAAAACAAATTGTCAATTATTACAATCCCAAAGGCGAATGACGAACTAAAATCT	64
CAGCAAAAAAAAAACAAGACAGGAAGAACGGTGGTGGAGTTGAGTAGAGAGAGGCCAGCAAAGC	64
CAGCAAAAAAAAAACCCTCTGCTCTGAAGGTGGAAGGTAGGATAATGCACCTGGATATCAGAAA	64 281
CAGCTTACCCTTGTTTGGTTGCCCGCATAACGCTTTATGCAAGTTTGCACAGTGCTTAAAGCAG	64 171
CTGCAACAACAAGCTCGTCGGCGCCAAGTTCTTCGGCCTGGGGTACGAGGCCGCGCACGGCGGG	64 290
CTGCAACAACGTGTCCTTCGCGCCATATGGCGGGAAGTGGCGCCGGGGCAAGAAGATCGCGGTG	64 382
CTGCAACAAGACAGAATTTGTTAAATGTAATATGAAACAGACACATCAAGTTAAATTCTGGTA	64 239
CTGCAACAAGCAGAAGGTCCGGCGGGGCTGTGGTCGCCGGAGGAGGACGAGAAGCTCATCAAG	64 197
CTGCAACAAGGAAATTGCAGGCGCGCGGATAACAATGTACCAATACAACGAAGATGCTCTCTGC	64 125
CTGCAACAAGTTTTTTGCGTGGGCAAGACTTCCATGCTTCTAGAGCCTGATCTCTCAAATCTC	64 141

Tags

Reads

MergeTagsByTaxa

Tags

```
CTGCCCTCGTTTAAGTCCGATTGCTGATAGTTATCTTTGTTTCTTACCAAATCCACATGGCTG  
CAGCAGTTGACTGGACATCTAGGGGCAAAGCACTGTTTCGGTGCGGGCTGAGATCGGAAGAGCG  
CAGCAGGCCGTGGTACATGTAATGGAGCATGGATTGAAGGTGGATGCCTTCATGTCCTGGACG
```



```
ATGCCTCTGCCCTCGTTTAAGTCCGATTGCTGATAGTTATCTTTGTTTCTTACCAAATCCACATGGCTGAGATCGGAAGAGCGGT  
GCGTCTGCCCTCGTTTAAGTCCGATTGCTGATAGTTATCTTTGTTTCTTACCAAATCCACATGGCTGAGATCGGAAGAGCGGT  
CGATCAGCAGTTGACTGGACATCTAGGGGCAAAGCACTGTTTCGGTGCGGGCTGAGATCGGAAGAGCGGTTTCAGCAGGAATGCCGA  
CCGGATATCAGCAGGCCGTGGTACATGTAATGGAGCATGGATTGAAGGTGGATGCCTTCATGTCCTGGACGCGGATCGAGCTCGC
```

SeqToTBTHDF5Plugin

Qseq file

Tags file

key file

TagsByTaxa (TBT)

(Presence/Absence matrix)

ATGCCT	BTx623
CGAT	N13
CCGGATAT	E36-1
CAGA	ICSL1118
AACT	ICSL1126
GCGT	ICSL1134

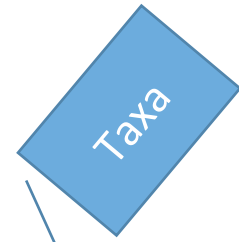
MergeTBTFiles

Merge TBT files

TBT file

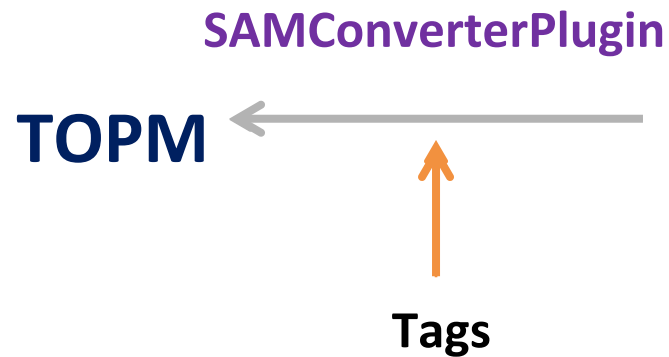
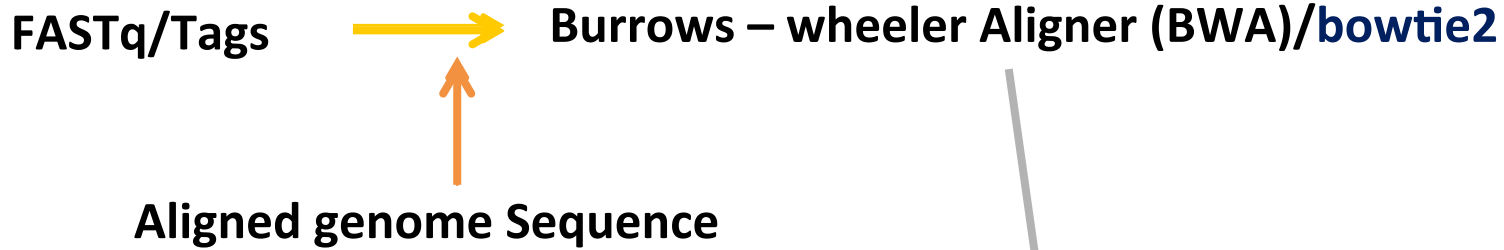
Tags

CTGCCCTCGTTTAAGTCCGATTGCTGATAGTTATCTTTGTTTCTTACCAAAATCCACATGGCTG
CAGCAAAAAAAAAAAGGACATGGGTATCCGGTAACACTGCAGAAAATTTGAGTAAAAGGCGGTA
CAGCAAAAAAAAAAATCTCAGAAAATGCAGTGCAGAGTGTATTTTTGCTTGATCCTGGCTCAC
CAGCAAAAAAAAAAACAAAAACAAAACAAAAGGATTTCTTTAGGGAAAAAACATAAGGTGCG
CAGCAAAAAAAAAACAATTGTCAATTATTACAATCCCAAAGGCGAATGACGAAACTAAAATCT
CAGCAAAAAAAAAACAAGACAGGAAGAACGGTGGTGGAGTTGAGTAGAGAGAGGCCAGCAAAGC
CAGCAAAAAAAAAACCCTCTGCTCTGAAGGTGGAAGGTAGGATAATGCACCTGGATATCAGAAA
CAGCTTACCCTTGTGGTTGCCCGCATACGCCTTTATGCAAGTTTGCACAGTGCTTAAAGCAG
CTGCAACAACAAGCTCGTCGGCGCCAAGTTCTTCGGCCTGGGGTACGAGGCCGCGCACGGCGGG
CTGCAACAACGTGTCCTTCGCGCCATATGGCGGGAAGTGGCGCCGGGGCAAGAAGATCGCGGTG
CTGCAACAAGACAGAATTTGTTAAATGTAATATGAAACAGACACATCAAGTTAAATTCTGGTA
CTGCAACAAGCAGAAGGTCCGGCGGGGCTGTGGTCCGGGAGGAGGACGAGAAGCTCATCAAG
CTGCAACAAGGAAATTGCAGGCGCGCGGATAACAATGTACCAATACAACGAAGATGCTCTCTGC
CTGCAACAAGTTTTTGCCTGGGCAAGACTTCCATGCTTCTAGAGCCTGATCTCTCAAAATCTC



	BTx623	N13	E36-1	ICSL1118	ICSL1126	ICSL1134
64	1	0	0	0	0	1
64	0	0	0	0	0	0
64	0	0	1	0	0	0
64	1	1	0	1	0	0
64	0	0	0	0	0	0
64	1	0	0	0	0	1
64	0	0	0	0	0	0
64	0	0	0	1	0	0
64	0	0	0	0	0	1
64	0	0	0	0	0	0
64	0	0	0	0	0	0
64	1	1	1	0	0	1
64	0	0	0	0	0	0
64	0	0	0	0	0	0

TagsOnPhysicalMap (TOPM)



Col	Field	Description
1	QNAME	Query (pair) NAME
2	FLAG	bitwise FLAG
3	RNAME	Reference sequence NAME
4	POS	1-based leftmost POSition/coordinate of clipped sequence
5	MAPQ	MAPping Quality (Phred-scaled)
6	CIAGR	extended CIGAR string
7	MRNM	Mate Reference sequence NaMe ('=' if same as RNAME)
8	MPOS	1-based Mate POSition
9	ISIZE	Inferred insert SIZE
10	SEQ	query SEquence on the same strand as the reference
11	QUAL	query QUALity (ASCII-33 gives the Phred base quality)
12	OPT	variable OPTional fields in the format TAG:VTYPE:VALUE

TOPM file

Tag	length	chr	start	end
CACCCCGCCAAGTGTCTGTGTTGAGGAGTCAACAAGAACCAATTGCAGGATGTAAAAGCAG	64	1	12982	13046
CTGCTGGTTAATACGTACTCCAGTTGGTTACCATGCACCAAGCAGGGCAAGGCACTCCTGCTCC	64	1	2132	2196
CTGCTGGTTAATACGTACTCCAGTTGGTTACTATGCACCAAGCAGGGCAAGGCACTCCTGCTCC	64	1	2132	2196
CAGCCCTACGGCCGCAACTGCAAAAATCAGATGTGAGCATAGTTCTTAATGTATCATAAAAAGT	64	2	2266316	2266380
CTGCTACCTCTGCCCTGCGTGCGTGACCCATACATACGGTGGGGCCGGGGTGTAATAATCTTGTG	64	2	2926106	2926170
CAGCGGCTGAACACTGACCAGTTGACGGCGTCGCGCGCGGGCAGGGCGACGATGTCGGCGCAG	64	3	2129422	2129486
CTGCGGCGGGCTCCTACTCCCTCAACGGAAGAGGGCGTTACGACGGGTGGAGGGCAGTTTCTCT	64	3	2140833	2140897
CTGCGGCGGGCTCCTACTCCCTCAACGGAAGAGGGCGTTACGACGGGGGGAGGGCAGTTTCTCT	64	3	2140833	2140897
CTGCGGCGGGCTCCTACTCCCTCAACGGAAGAGGGCGTTACGACGGGGGGAGGGCAGTTTCTCT	64	3	2140833	2140897
CTGCGGCGGGCTCCTACTCCCTCAACGGAAGAGGGCGTTACGACGGGGGGAGTGGCAGTTTCTCT	64	3	2140833	2140897
CTGCGGCGGGCTCCTACTCCCTCAACGGAAGAGGGCGTTACGACGGGGGGGGGGCAGTTTCTCT	64	3	2140833	2140897
CTGCGGCGGGCTCCTACTCCCTCAACGGAAGAGGGCGTTACGACGGGGGGGGTGGCAGTTTCTCT	64	3	2140833	2140897
CTGCGGCGGGCTCCTACTCCCTCAACGGAAGAGGGCGTTACGACGGGTGGAGGGCAGTTTCTCT	64	3	2140833	2140897
CTGCGGCGGGCTCCTACTCCCTCAACGGAAGAGGGCGTTACGACGGGTGGAGTGGCAGTTTCTCT	64	3	2140833	2140897
CAGCTGTCCCTTGTAACCTGCGGACATGACCGAACATCCAAGTACTCGAGAGCCTTGAAGGCTT	64	4	56916960	56917024
CTGCTTGTGTTGTTGATTGCTCGTTGGCCGATCGGTGGTCGTTGTCGTCGTTGAGCTGGTAGC	64	4	56942951	56943015
CAGCGCCCAAACCTTCTTTCCCGCTCTTGTTGATTCTCAGTCTCTCAGTTCTTGAACATCG	64	5	54508281	54508345
CTGCCTCAGCACAGCCGGTGCTCTTGGATCCTCGAGCACCTGCACTCTTTCAGGAGGCCAAGGA	64	5	13747204	13747268
CAGCACGCGTATCCTTGGCCCTTATTGTTTGAAAATGACCTGTGACGGTGCCTTCTGCGTAGT	64	6	52338128	52338192
CTGCGGCGGGCATGACCGACGGGCGGCGTGGGTGCCGAGTGCCGGCAACTCCGCCGTTCT	64	6	58276481	58276545
CTGCCACAGACGGAGTTACGACAACCTTGCAAACTCCGCCCGGCTTAAGTTAATTTAATCAGG	64	7	18492751	18492815
CAGCCAAAAAATGGCAAAAGGTGTAGGACTTGGGTACGTCAAAATTCATTCATGTGAGAATGA	64	7	19393217	19393281

Hapmap file

Key file
Qseq file
Tags file
TOPM file
TBT file

TagsToSNPByAlignmentPlugin

Hapmap file
(Sequence alignment file)

```
CTGCGGC GGGCTCC TACTCCC TCAACGGAAGAGGCGTTACGACGGGGGGAGTGGCAGTTT
CTGCGGC GGGCTCC TACTCCC TCAACGGAAGAGGCGTTACGACGGGGGGGGTGGCAGGTT
CTGCGGC GGGCTCC TACTCCC TCAACGGAAGAGGCGTTACGACGGGGGGGGGGCAGGTT
CTGCGGC GGGCTCC TACTCCC TCAACGGAAGAGGCGTTACGACGGGGGGAGTGGCAGGTT
CTGCGGC GGGCTCC TACTCCC TCAACGGAAGAGGCGTTACGACGGGTGGAGTGGCAGTTT
CTGCGGC GGGCTCC TACTCCC TCAACGGAAGAGGCGTTACGACGGGTGGAGTGGCAGGTT
CTGCGGC GGGCTCC TACTCCC TCAACGGAAGAGGCGTTACGACGGGGGGAGGGGCAGGTT
CTGCGGC GGGCTCC TACTCCC TCAACGGAAGAGGCGTTACGACGGGTGGAGGGGCAGTTT
CTGCGGC GGGCTCC TACTCCC TCAACGGAAGAGGCGTTACGACGGGGGGAGGGGCAGTTT
*****
```



Why production pipeline?

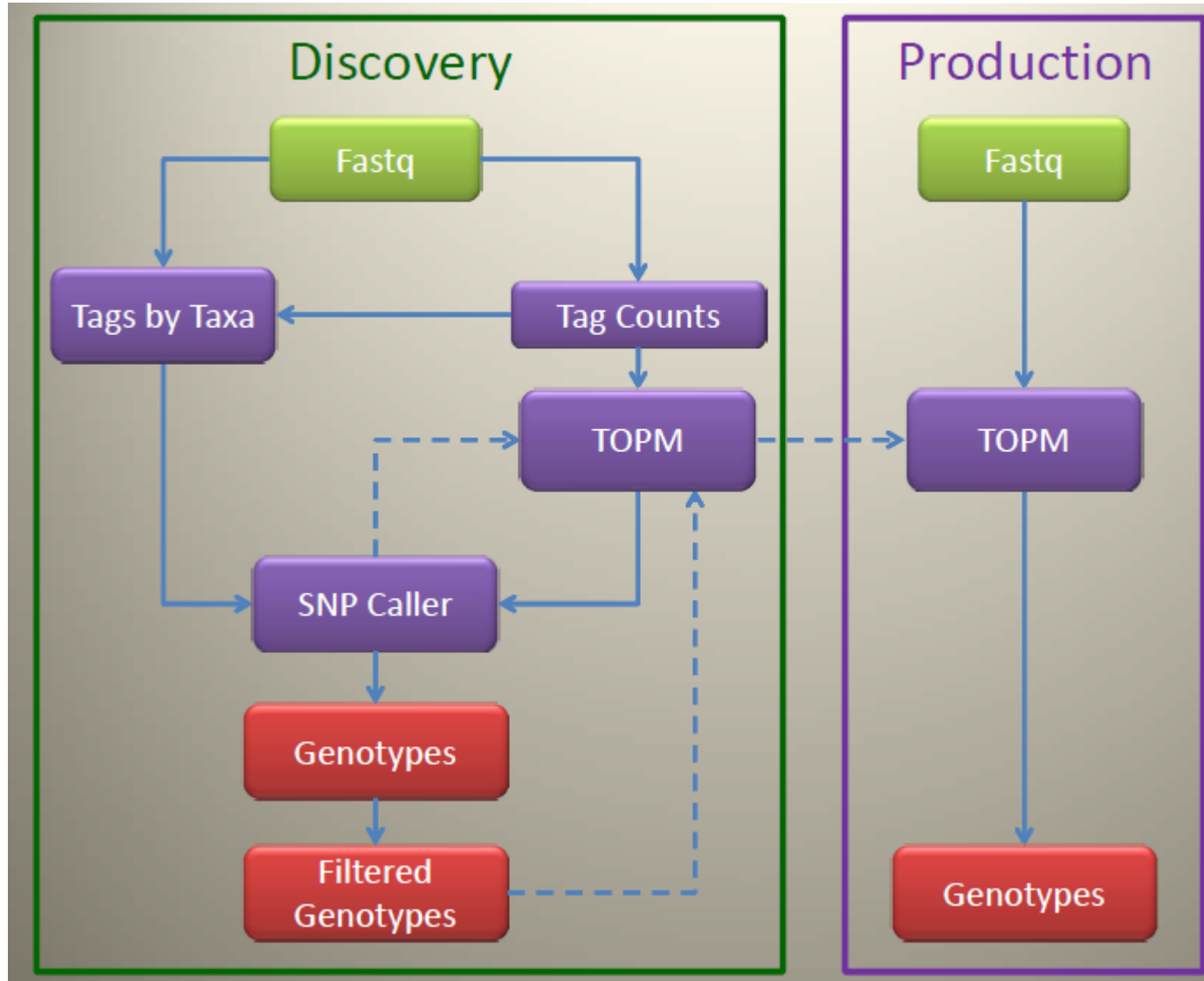
- ❖ **Discovery pipeline takes more time**

 - Eg: Last maize build took > 3 months**

- ❖ **Most common alleles have been identified in discovery pipelines**

- ❖ **Use of available information from discovery pipelines to call SNPs
in new run**

GBS Production pipeline



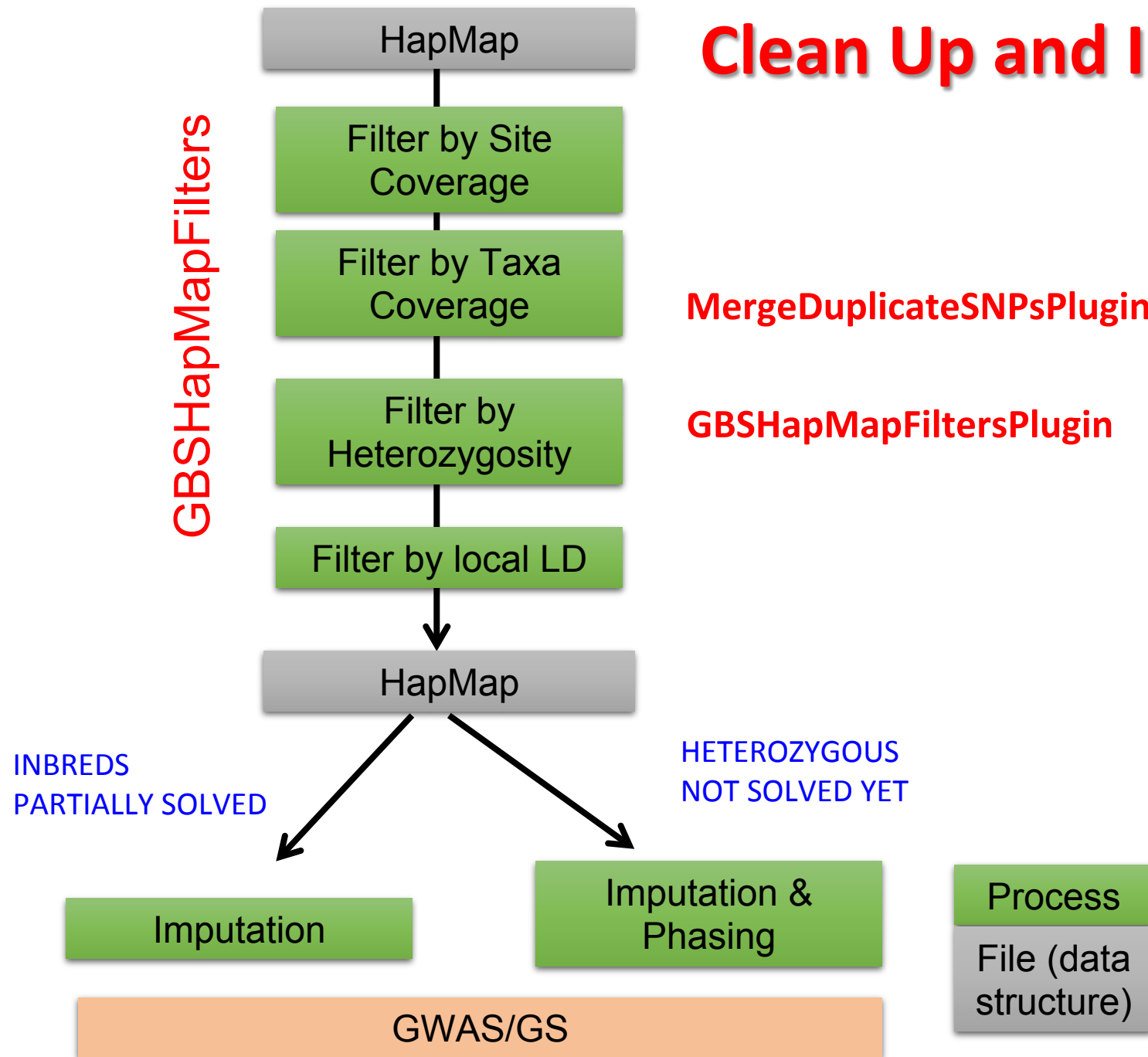
UNEAK pipeline

(Universal Network Enabled Analysis Kit)
A reference free SNP calling pipeline

Designed for species that.....

- lack a reference genome
- diploid or polyploid
- inbreeders or outcrosses
- having limited genetic or genomic resources

Clean Up and Imputation



Hapmap file

rs#	alleles	chrom	pos	ICSB617	ICSB212	ICSB536	ICSB79	ICSB538	ICSB639	ICSB417	ICSB748	ICSB563
S9_20491	G/A	9	20491	G	N	N	N	N	N	N	N	N
S9_23432	C/Y	9	23432	N	N	Y	C	N	N	C	C	N
S9_23452	T/K	9	23452	N	N	T	T	N	N	T	T	N
S9_26213	G/A	9	26213	N	N	N	W	N	N	N	G	N
S9_27701	T/C	9	27701	T	N	T	N	N	Y	C	T	N
S9_27720	T/K	9	27720	T	N	T	N	N	T	T	T	N
S9_30044	C/T	9	30044	C	N	N	N	N	N	N	N	N
S9_36986	A/G	9	36986	N	N	A	N	N	N	N	N	N
S9_39902	C/T	9	39902	N	N	N	N	T	N	T	N	T
S9_39912	A/G	9	39912	N	N	N	N	G	N	G	N	G
S9_40450	A/C	9	40450	N	K	N	N	N	N	N	N	N
S9_40452	G/T	9	40452	N	N	N	N	N	N	N	N	N
S9_40459	A/G	9	40459	N	N	N	N	N	N	N	N	N
S9_43958	T/C	9	43958	N	N	N	Y	N	M	N	N	N
S9_43982	C/T	9	43982	N	N	N	N	N	N	N	N	N
S9_44191	G/C	9	44191	N	M	N	N	N	N	N	N	N
S9_45394	T/A	9	45394	N	N	N	N	N	N	N	N	N
S9_56733	C/M	9	56733	N	N	N	N	N	N	C	N	N
S9_56734	C/S	9	56734	N	N	N	N	N	N	C	N	N
S9_56735	T/W	9	56735	N	N	N	N	N	N	T	N	N
S9_69001	C/Y	9	69001	C	N	N	N	N	C	N	C	N
S9_72163	C/T	9	72163	N	N	N	N	N	N	T	N	N
S9_74270	C/T	9	74270	C	N	T	N	N	N	C	N	N
S9_76335	A/G	9	76335	N	A	N	N	N	N	R	A	N
S9_76341	T/G	9	76341	N	T	N	N	N	N	K	T	N
S9_78501	A/C	9	78501	N	A	N	N	N	N	N	N	N

Hapmap Files

sorg_Jan2013_Filter2_c10.hmp.txt	1001,347,276	Text Doc...
sorg_Jan2013_Filter2_c9.hmp.txt	972,503,003	Text Doc...
sorg_Jan2013_Filter2_c8.hmp.txt	667,513,134	Text Doc...
sorg_Jan2013_Filter2_c7.hmp.txt	616,327,520	Text Doc...
sorg_Jan2013_Filter2_c6.hmp.txt	1048,787,660	Text Doc...
sorg_Jan2013_Filter2_c5.hmp.txt	1047,007,421	Text Doc...
sorg_Jan2013_Filter2_c4.hmp.txt	1284,135,657	Text Doc...
sorg_Jan2013_Filter2_c3.hmp.txt	1505,358,845	Text Doc...
sorg_Jan2013_Filter2_c2.hmp.txt	1493,688,785	Text Doc...
sorg_Jan2013_Filter2_c1.hmp.txt	1675,855,191	Text Doc...

Chromosome	Unfilter	Filter2 (mnMAF 0.01, mnTCov 0.01, mnSCov 0.01, mnF 0.8)
1	1288506	83460
2	999490	74192
3	1077534	74976
4	868854	64199
5	528132	52401
6	706426	51633
7	569982	28153
8	462068	30221
9	618291	48143
10	659650	49595
Total	7778933	556973

Optimizing GBS in New Species



A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species

Robert J. Elshire¹, Jeffrey C. Glaubitz¹, Qi Sun², Jesse A. Poland³, Ken Kawamoto¹, Edward S. Buckler^{1,4}, Sharon E. Mitchell^{1*}

1 Institute for Genomic Diversity, Cornell University, Ithaca, New York, United States of America, **2** Computational Biology Service Unit, Cornell University, Ithaca, New York, United States of America, **3** Hard Winter Wheat Genetics Research Unit, United States Department of Agriculture/Agricultural Research Service, Manhattan, Kansas, United States of America, **4** Plant, Soil and Nutrition Research Unit, United States Department of Agriculture/Agricultural Research Service, Ithaca, New York, United States of America

Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol

Fei Lu¹, Alexander E. Lipka^{1,2}, Jeff Glaubitz¹, Rob Elshire¹, Jerome H. Cherney³, Michael D. Casler^{4,5}, Edward S. Buckler^{1,2}, Denise E. Costich^{1,2*}

1 Institute for Genomic Diversity, Cornell University, Ithaca, New York, United States of America, **2** Agricultural Research Service, United States Department of Agriculture, Ithaca, New York, United States of America, **3** Department of Crop and Soil Sciences, Cornell University, Ithaca, New York, United States of America, **4** Agricultural Research Service, United States Department of Agriculture, Madison, Wisconsin, United States of America, **5** Department of Agronomy, University of Wisconsin–Madison, Madison, Wisconsin, United States of America

Population genomic and genome-wide association studies of agroclimatic traits in sorghum

Geoffrey P. Morris^{a,1,2}, Punna Ramu^{b,1}, Santosh P. Deshpande^b, C. Thomas Hash^c, Trushar Shah^b, Hari D. Upadhyaya^b, Oscar Riera-Lizarazu^b, Patrick J. Brown^d, Charlotte B. Acharya^e, Sharon E. Mitchell^e, James Harriman^e, Jeffrey C. Glaubitz^e, Edward S. Buckler^{e,f,g}, and Stephen Kresovich^a

^aDepartment of Biological Sciences, University of South Carolina, Columbia, SC 29208; ^bInternational Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad 502 324, Andhra Pradesh, India; ^cICRISAT-Sadoré, BP 12404 Niamey, Niger; ^dDepartment of Crop Sciences, University of Illinois, Urbana, IL 61801; ^eInstitute for Genomic Diversity and ^fDepartment of Plant Breeding and Genetics, Cornell University, Ithaca, NY 14853; and ^gAgricultural Research Service, Department of Agriculture, Ithaca, NY 14853

Acknowledgements



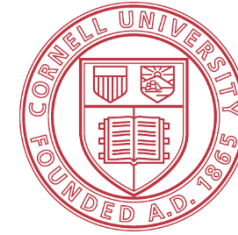
Odeny Damaris

Stefania Grando

Santosh Deshpande

Trushar Shah

Punna Ramu



Ed Buckler

Jeff Glaubitz

James Harriman

Terry C

Rob Elshire

Sharon E. Mitchell

Alex Lipka

Fei Lu

Thank You



for your kind attention