Using High Throughput Genotyping Information for Management of a Large Maize Collection

By

Candice Gardner, Maria Cinta Romay, Mark Millard, Sherry Flint-Garcia, James Holland, Edward Buckler

October, 2012 CSSA Meeting

Peru, 2008
The World's #1 Crop - 91 M U.S. Acres

The Other Members of the Top 10 World Crops

Wheat

Rice

Potatoes

Cassava

Soybean

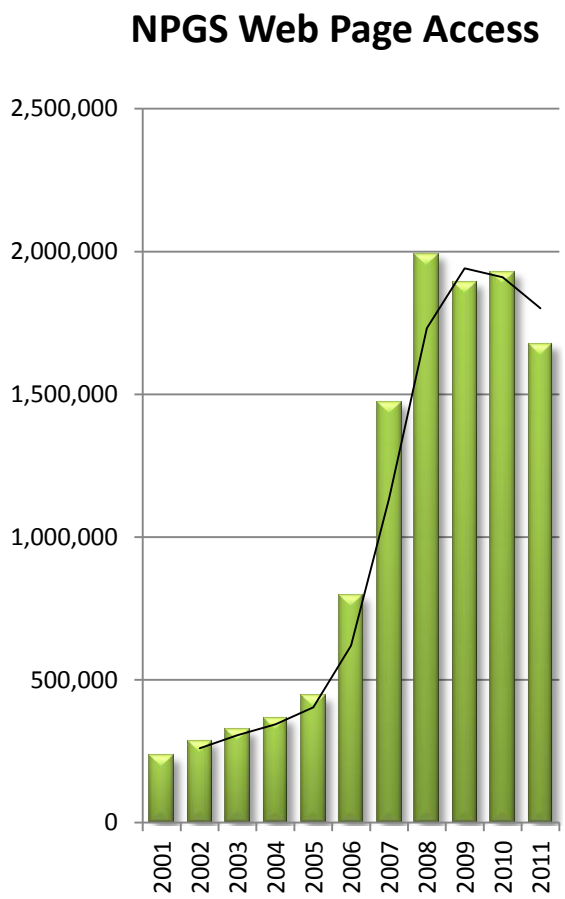
Sweet Potatoes

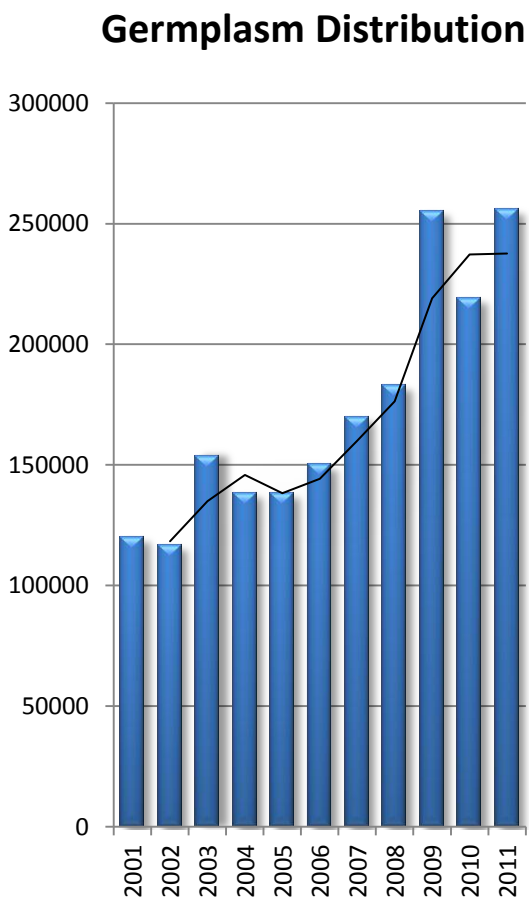
Sorghum

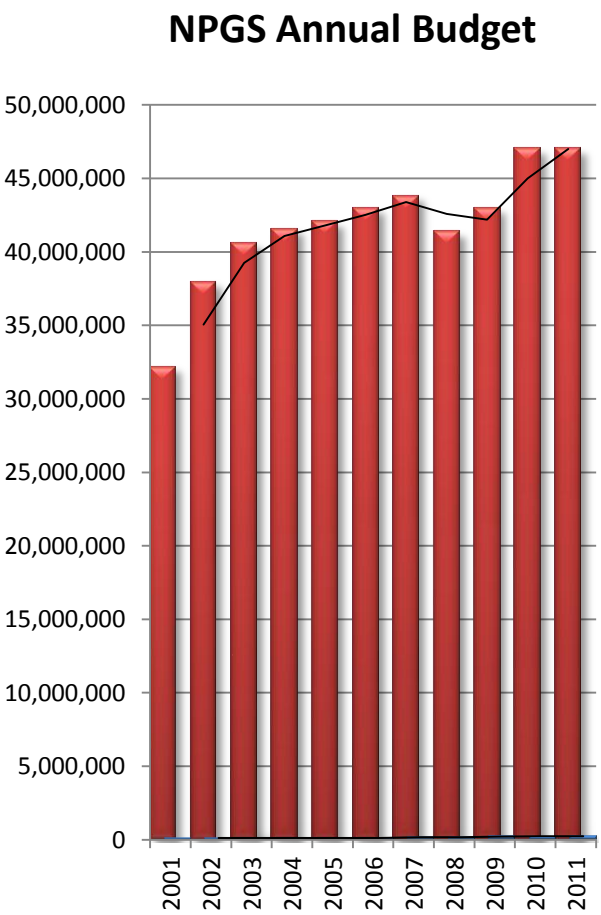
Yams

Plantains

Demand for NPGS germplasm and information vs. NPGS budget





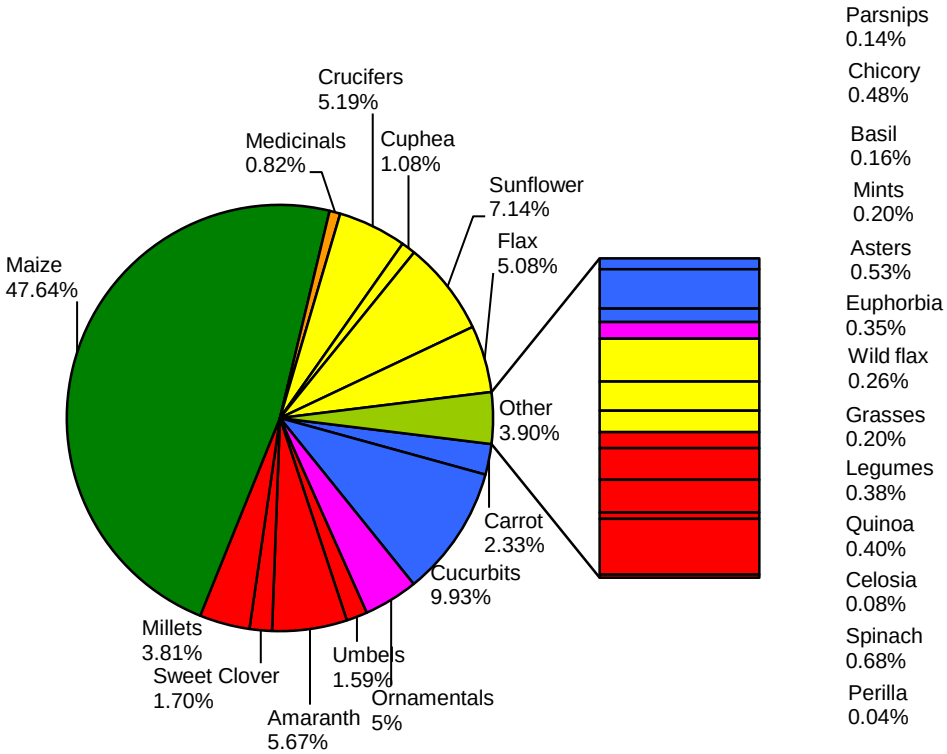


Courtesy of P.K. Bretting, 2012

Use of the Collections

- **Trait evaluation and extraction**
  - ❖Morphological, agronomic, horticultural and quality
  - ❖Biochemical, molecular
  - ❖Aesthetic
- **Varietal improvement and crop development**
- **Industrial applications**
- **Genomic and genetic investigations**
- **Taxonomic research**
- **Other basic and applied research**
- **Education and outreach**
- **Repatriation**

NCRPIS Holdings – October, 2012
52,075 Accessions – 1,925 Taxa

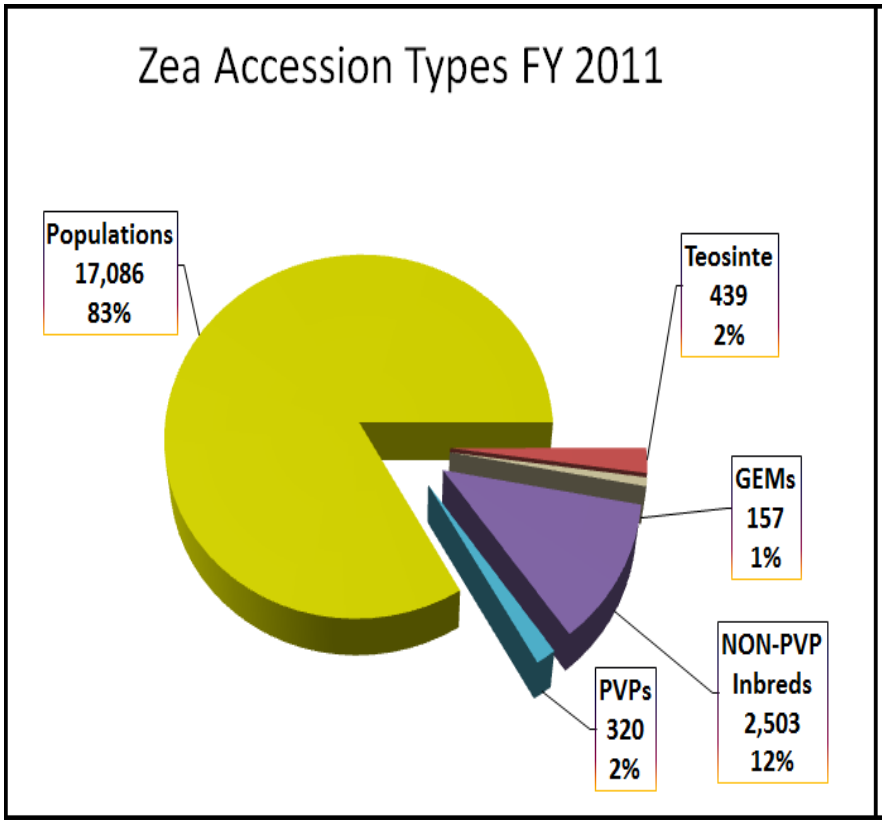


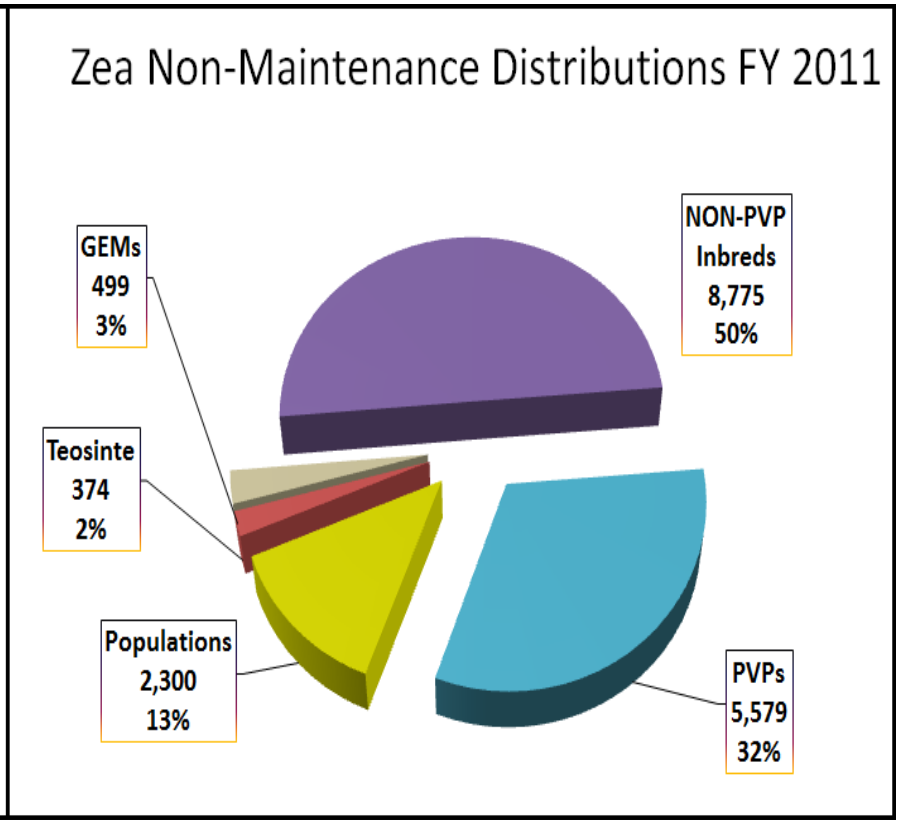
Plant Germplasm Accessions by NPGS Site *

Genebank Site	Count	% of NPGS Holdings	% of NPGS Distributions**	Avg Annual Distributions**
National Small Grains Collection	139,941	26	24.4	36,626
Plant Genetic Resources Conservation Unit, Griffin, GA	91,843	17	17.7	26,508
Western Regional PI Station	89,144	16	16.4	24,687
North Central Regional PI Station	51,576	9	16.7	25,005
Total – 4 Sites	380,014	69	75.2	112,826
Total - All NPGS Sites (20)	543,863	100	100	150,095
* November, 2011 data – excludes genetic stocks data				
** 2005 - 2010 Data				
Genebank Site	Count	% of NPGS Holdings	% of NPGS Distributions**	Avg Annual Distributions**
Maize Genetic Stock Center	7,510	1	58.7	10, 186

Although the NCRPIS holds about 9% of NPGS accessions, it accounts for 16.7% of annual distributions. Of these, 40-50% are maize germplasm.

Maize Collection Holdings and Availability October 1, 2011 Status





Comparison of Collection and Distribution Groups

Researchers frequently want….

- Linked genomic and phenotypic information
- Ability to determine value of alleles, including cryptic alleles, haplotypes
- Access to standardized information
- Access to descendant germplasm populations

Key Challenges

- Fulfilling the demand for additional germplasm characterizations/evaluations
- Managing large collections
- Understanding the composition of the collections
- Facilitating utilization by providing useful quality information

Courtesy of P.K. Bretting, 2012

From this point through slide 35, most slides were developed by Cinta Romay for:
Maize inbred lines at the NCRPIS*: a great public resource to explore maize genetics

M. Cinta Romay; Mark Millard; Zhiwu Zhang; Jason A. Peiffer; Jeffrey C. Glaubitz; Sharon E. Mitchell; Sherry Flint-Garcia; Michael D. McMullen; James B. Holland; Edward S. Buckler; Candice Gardner

Maize meeting 2012

* PI Station in Ames, IA

Objectives:

- Evaluate the genetic diversity present at the entire USDA maize inbred lines collection
  - Confirm germplasm identity and help curational management
  - Identify additional materials that could be requested to improve the collection
- Understand population structure across the genome and subgroups of germplasm
- Evaluate the use of the collection and GBS markers for genome-wide association studies
- Facilitate targeted use of the collection

The germplasm: Ames inbred panel

TOTAL: 2,812 unique taxa (2,711 from the USDA collection in Ames, IA)

Includes different breeding materials:

- 213 ExPVP
- 212 GEM (Genetic Enhancement of Maize)
- 179 classified as Stiff Stalk
- 187 classified as Non-Stiff Stalk
- 141 sweet corn
- 80 popcorn
- 402 from tropical origin

Different breeding origins:

- 31 different countries (USA, Spain, Mexico, Thailand, China, Nigeria, etc.)
- 35 different US states
- Almost a century of breeding efforts

Phenotypic Data

Locations:

- 2010 – Geneva, NY, Columbia, MO, and Raleigh, NC locations (Ames destroyed by storms)

-2011 – Ames, IA (2 reps)

Traits from all locations:

GDUSHD, GDUSLK, ASI, Plant ht, Ear ht, Number of Nodes, Number of Nodes to Ear & Above Ear

Traits from Ames:

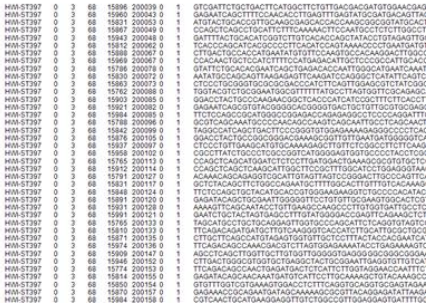
Shank length, Shank node #, # Husk Leaves, Husk Extension, # kernel rows, ear length, ear diameter, kernel row orientation, cob color, etc.

The tool: Genotyping By Sequencing (GBS)

Digest DNA with RE *Ape*KI 96-plex (Elshire et al. 2011)

Run on Illumina

Billions of good reads



Process through GBS reference genome pipeline

Filtered with GBS Hapmap filters (Site and Taxa coverage, Heterozygosity and local LD)

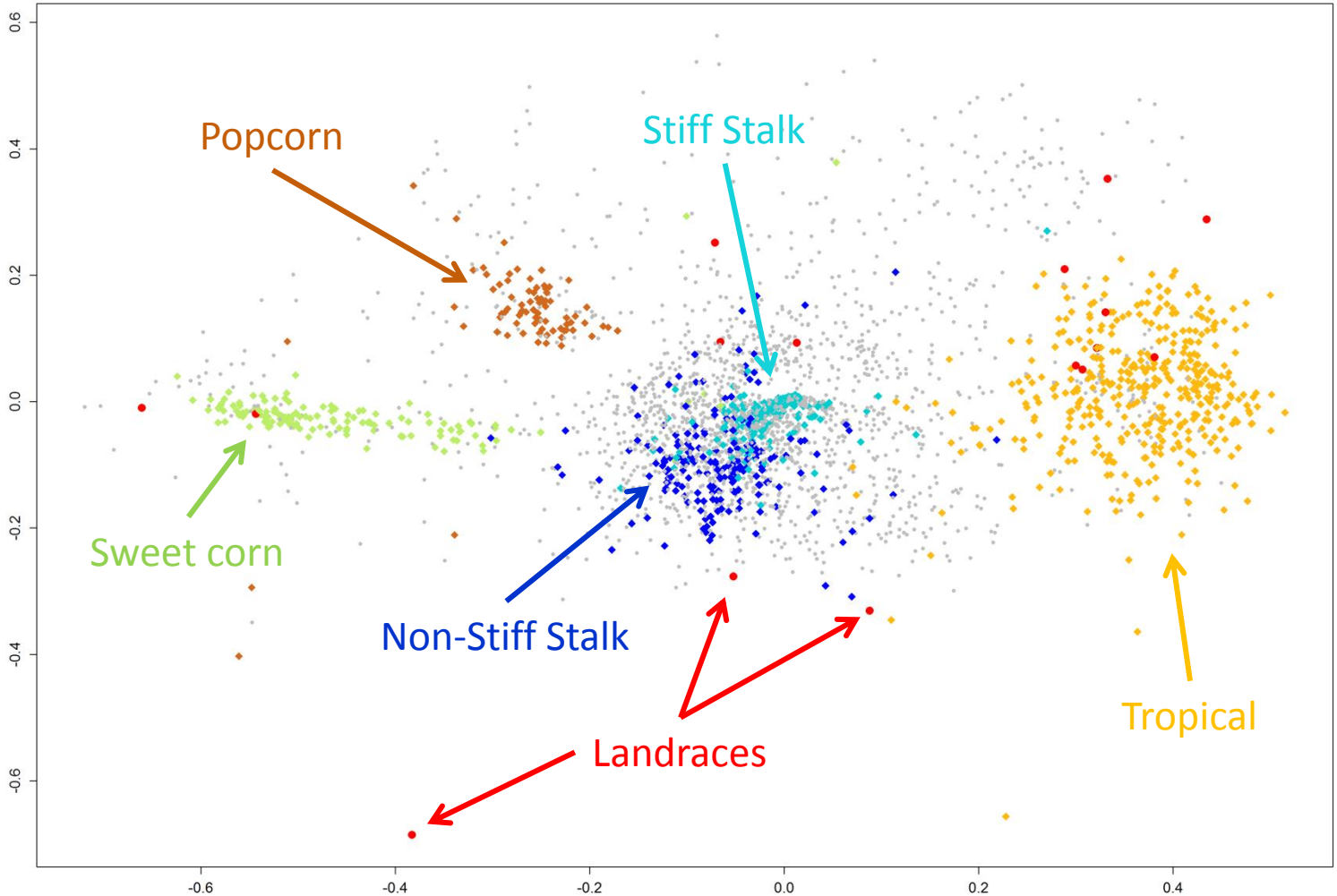
681,257 SNPs

- SNP discovery based on sample of ≈22,000 Zea taxa
- 14% of tags are too repetitive
- Millions of frequent tags genetically map

GBS discovers low frequency alleles that can be mapped

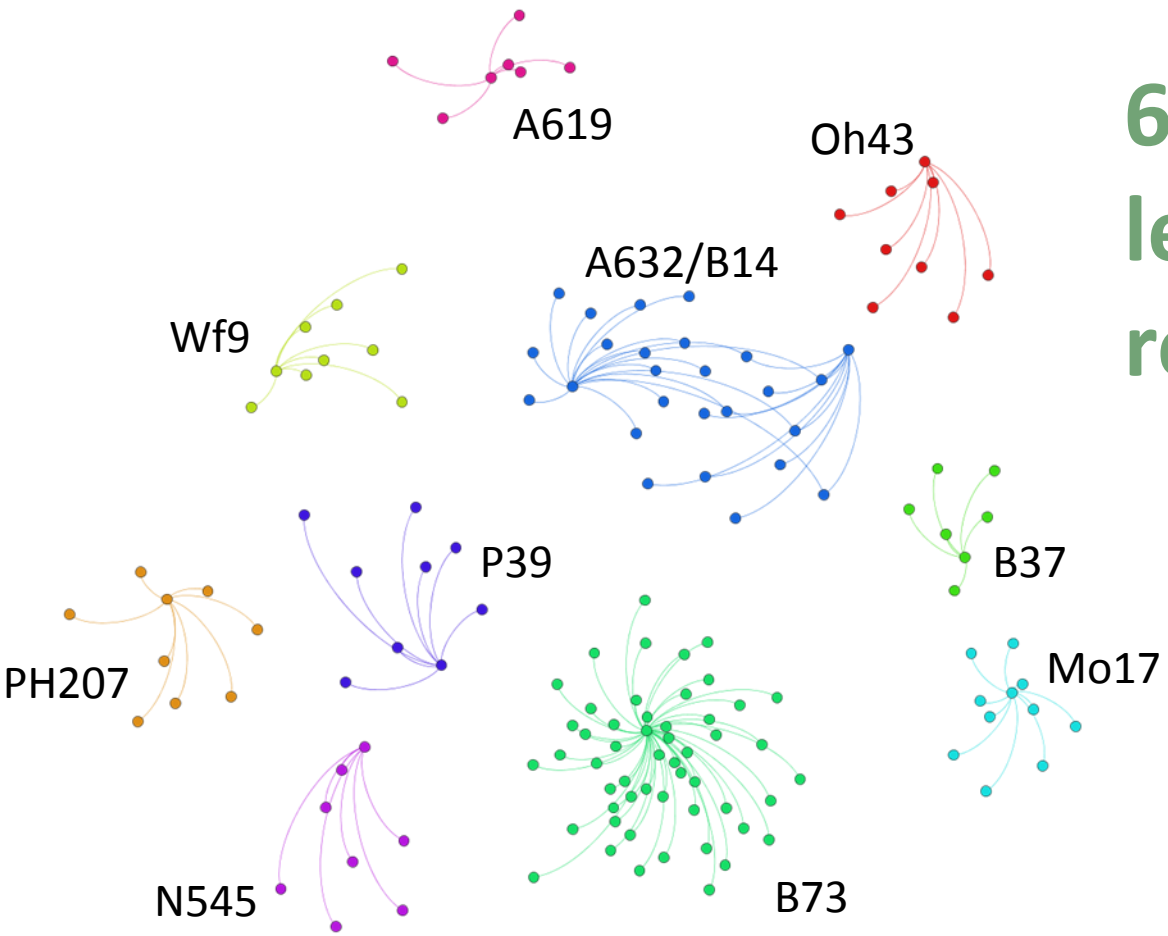
There is some B73 bias due to the use of reference genome

GBS explains the genetic variation among the 2,812 maize inbreds



Principal Coordinate Analysis of the maize inbreds using information from GBS markers

Some historical inbreds create clusters with more than 10 lines

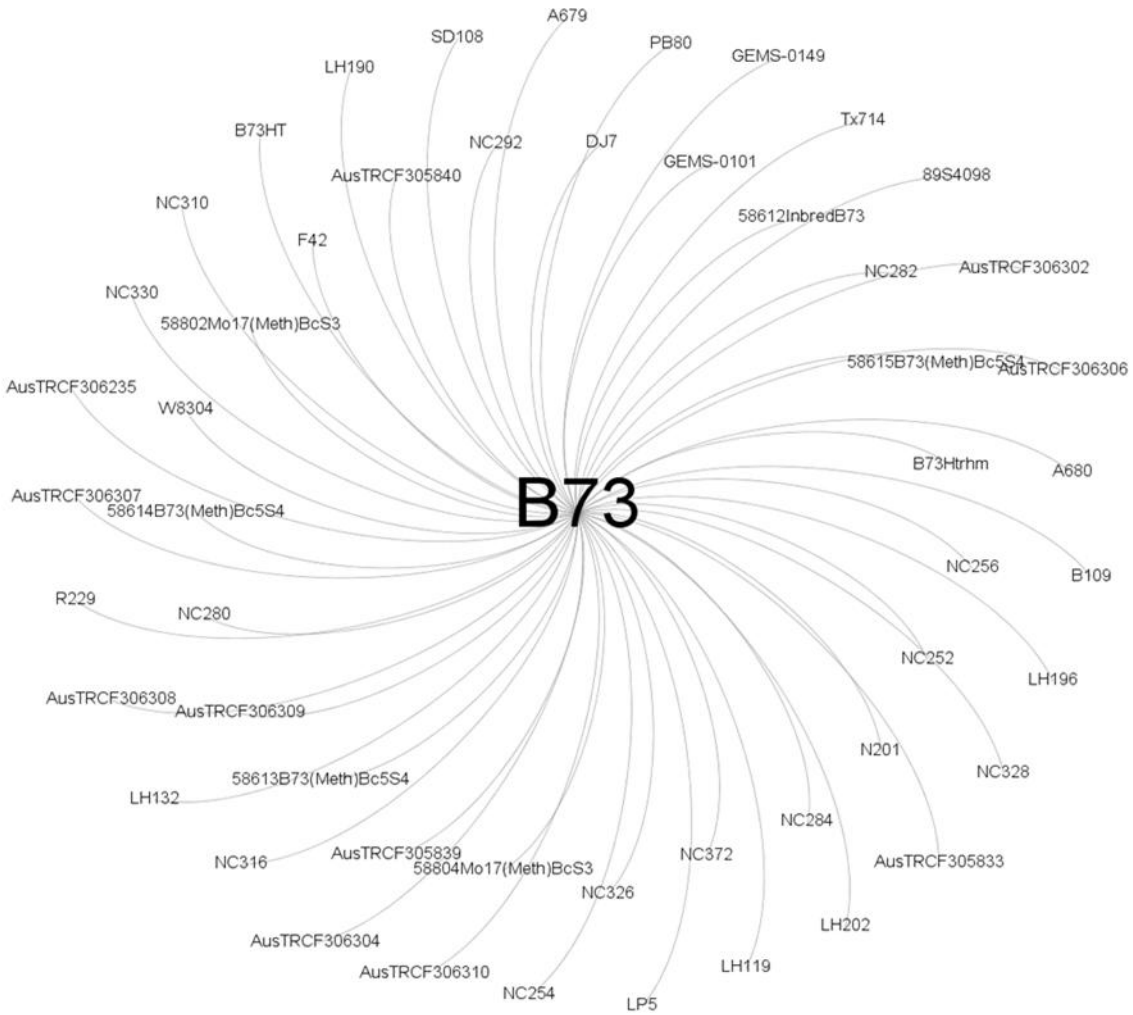


687 lines have at least one BC4-like relation

Network of inbreds more than 0.96 IBD

B73 is closely related to more than 50 other inbreds

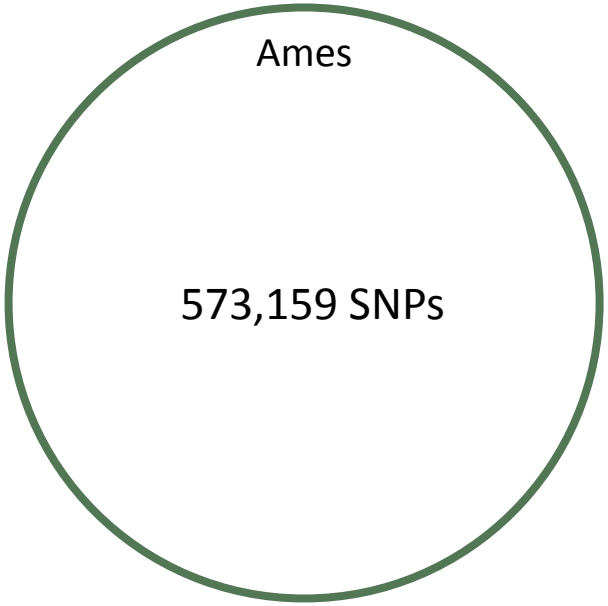
GBS relations can help identify errors

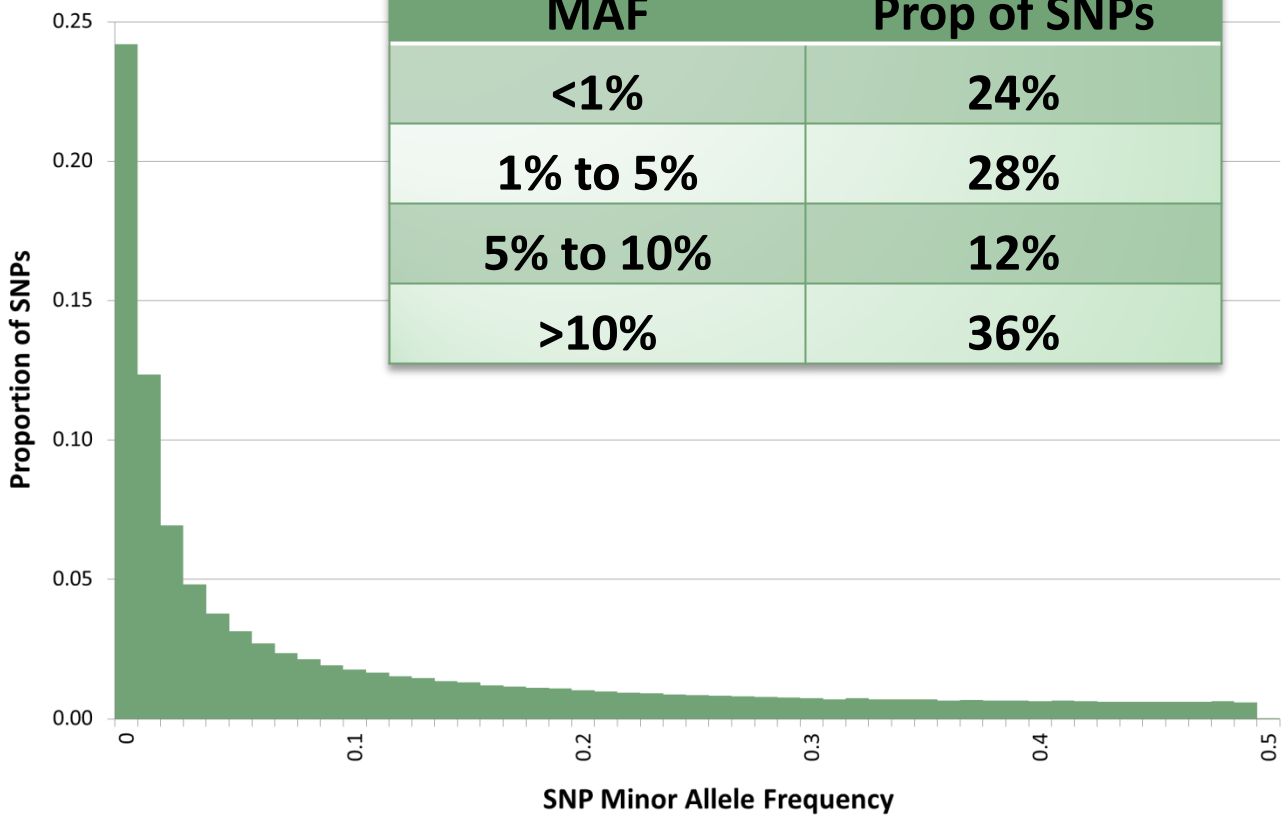


Network relations for inbreds more than 0.96 IBD with B73

How are alleles distributed?

Most SNPs are rare in the Ames inbred panel



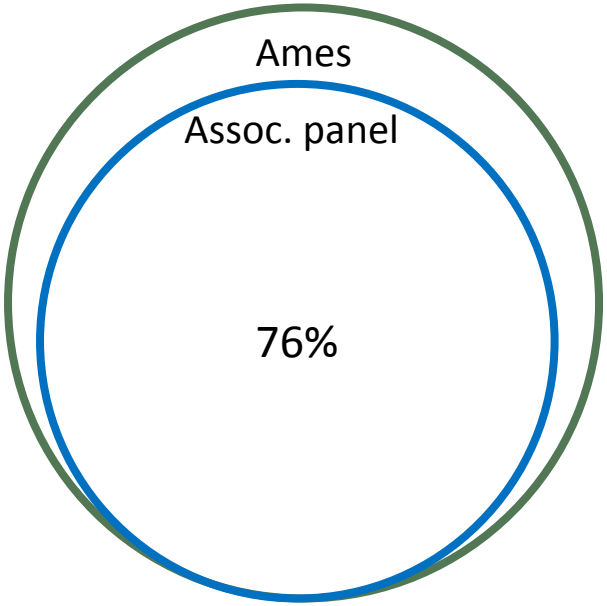


MAF	Prop of SNPs
<1%	24%
1% to 5%	28%
5% to 10%	12%
>10%	36%

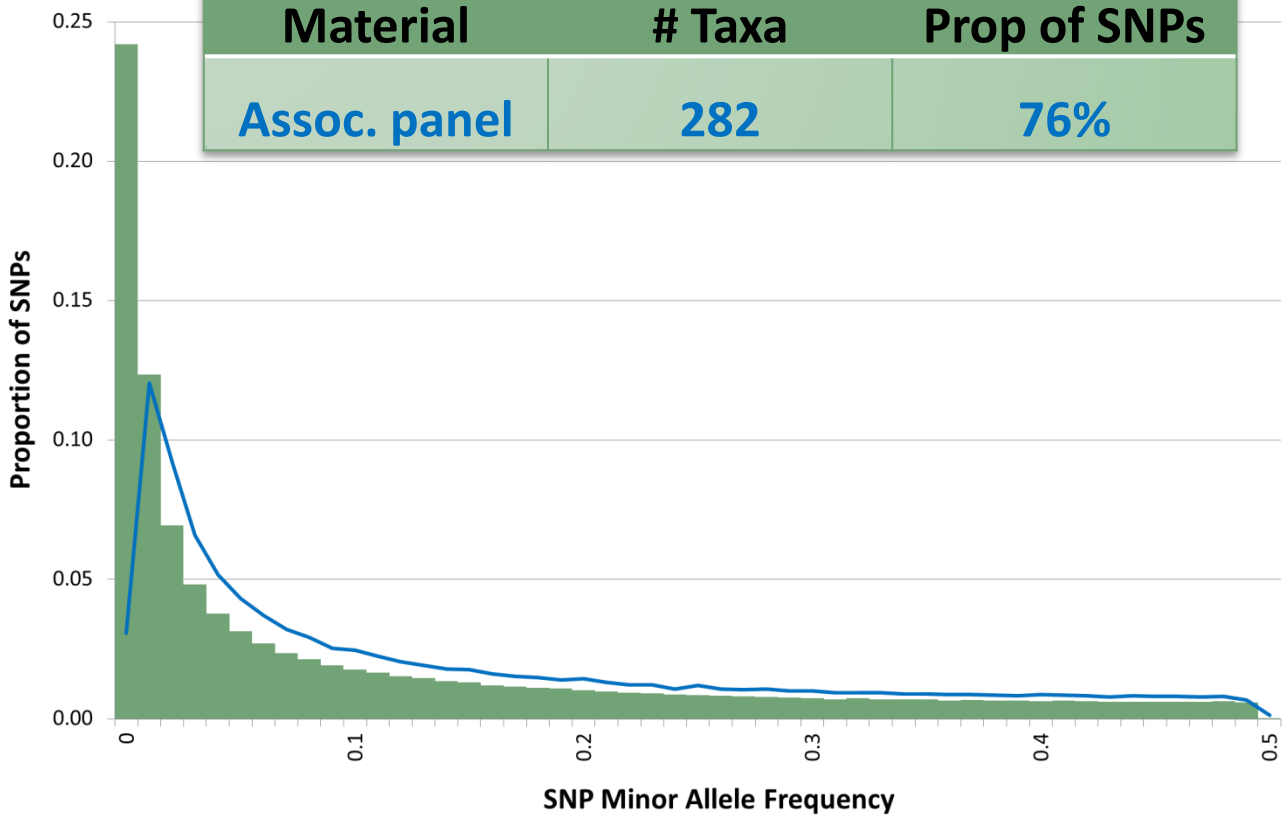
SNP distribution across 2,709 lines

How are alleles distributed?

Goodman association panel is a good subsample of the whole Ames panel



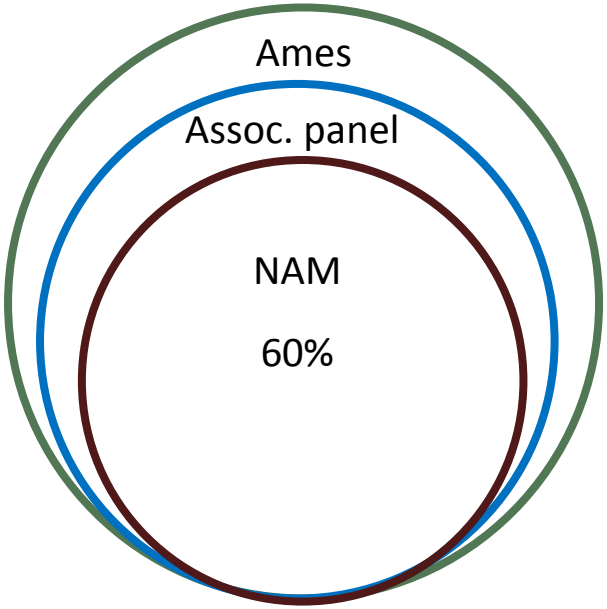
Material	# Taxa	Prop of SNPs
Assoc. panel	282	76%



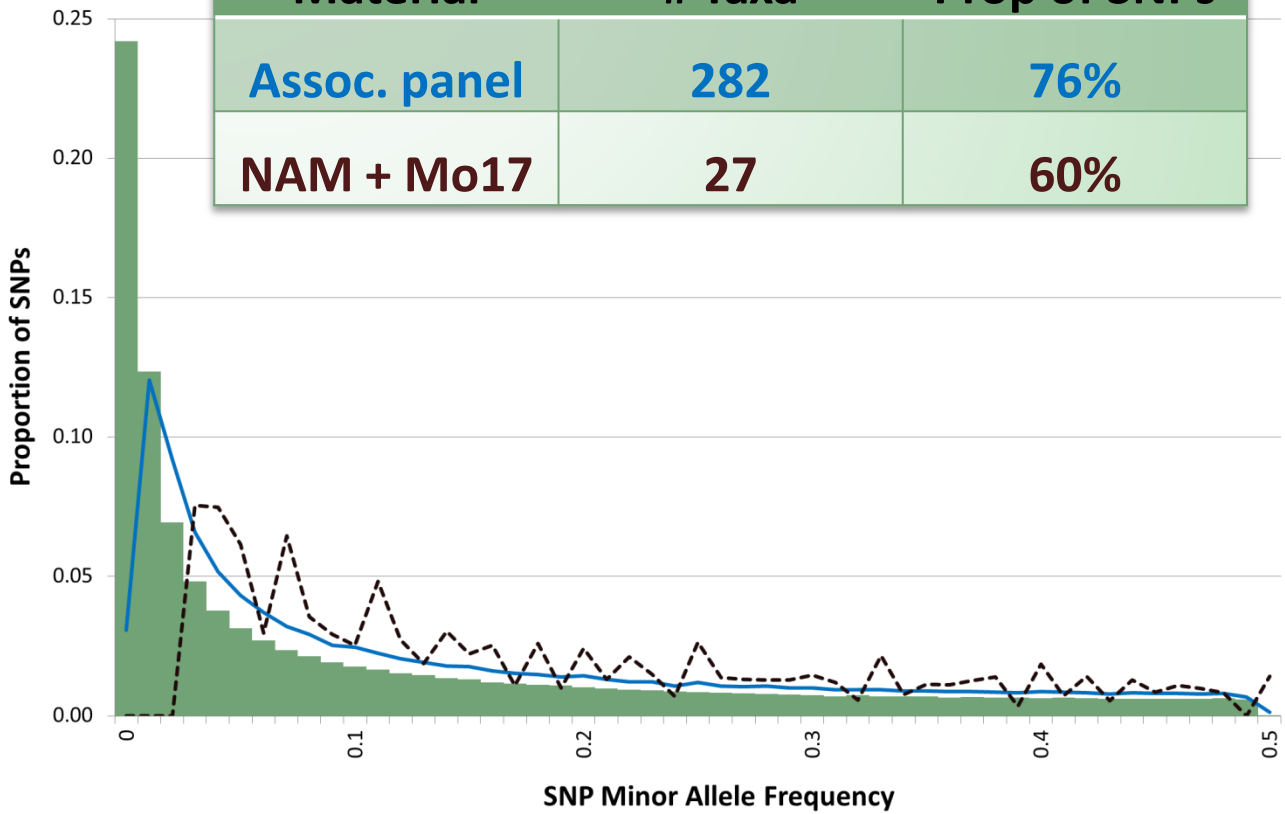
SNP distribution across 2,709 lines vs. 282 lines in the association panel

How are alleles distributed?

With a small sample NAM captures over a half of overall diversity



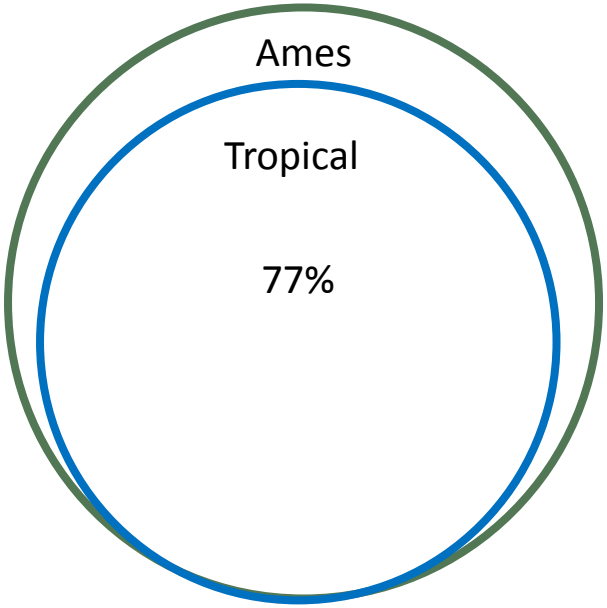
Material	# Taxa	Prop of SNPs
Assoc. panel	282	76%
NAM + Mo17	27	60%



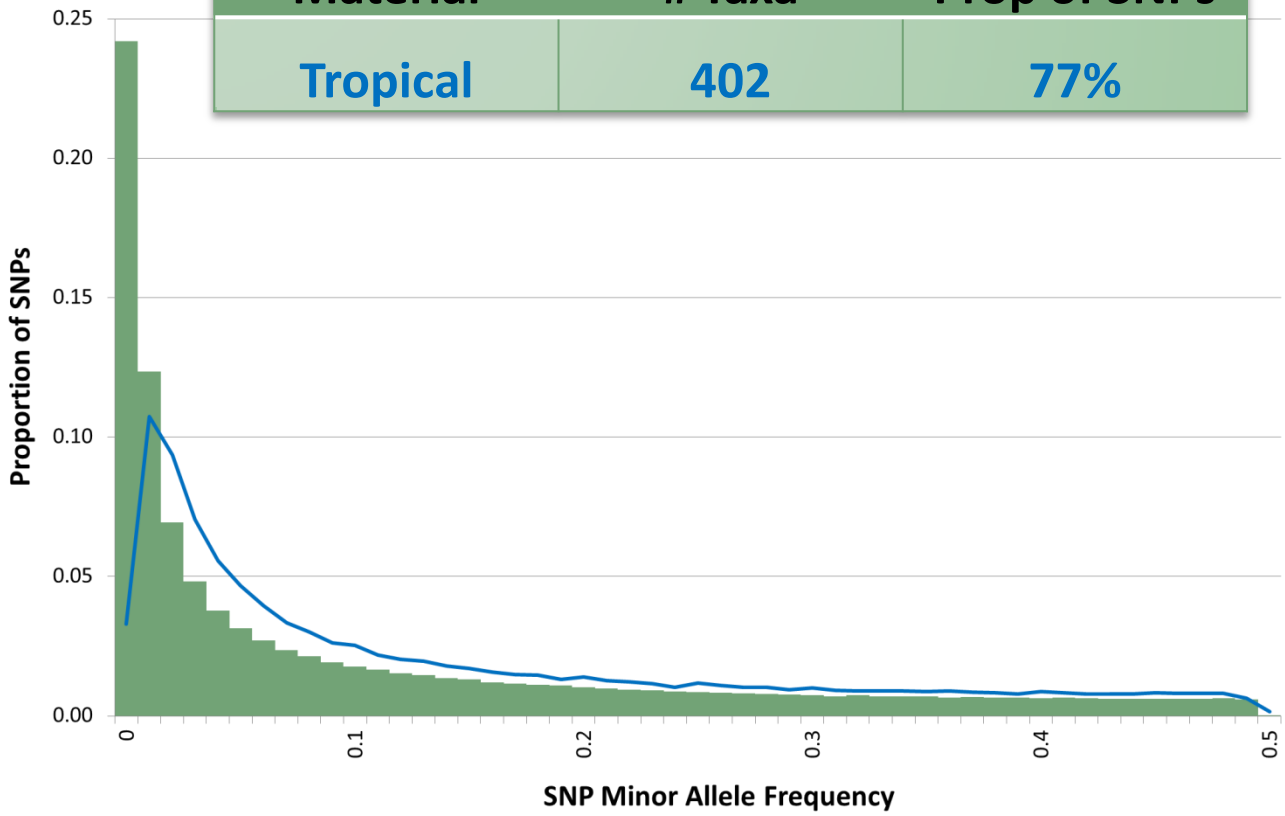
SNP distribution across 2,709 lines vs. 282 lines in the association panel and NAM parents

How are alleles distributed?

Tropical inbreds captures the most allelic diversity



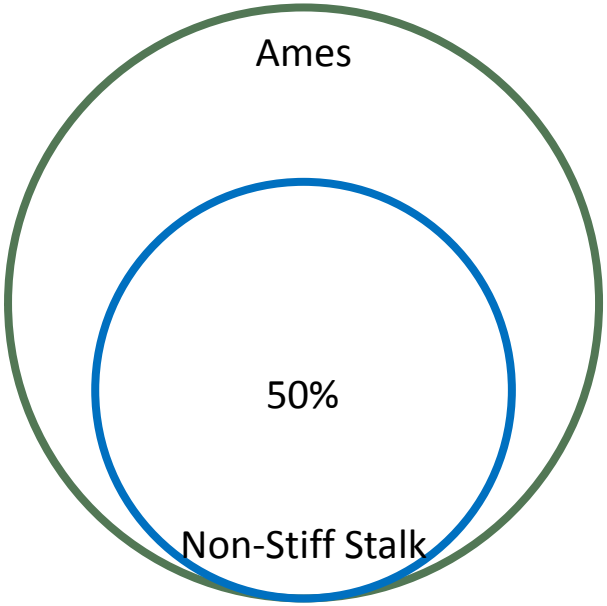
Material	# Taxa	Prop of SNPs
Tropical	402	77%



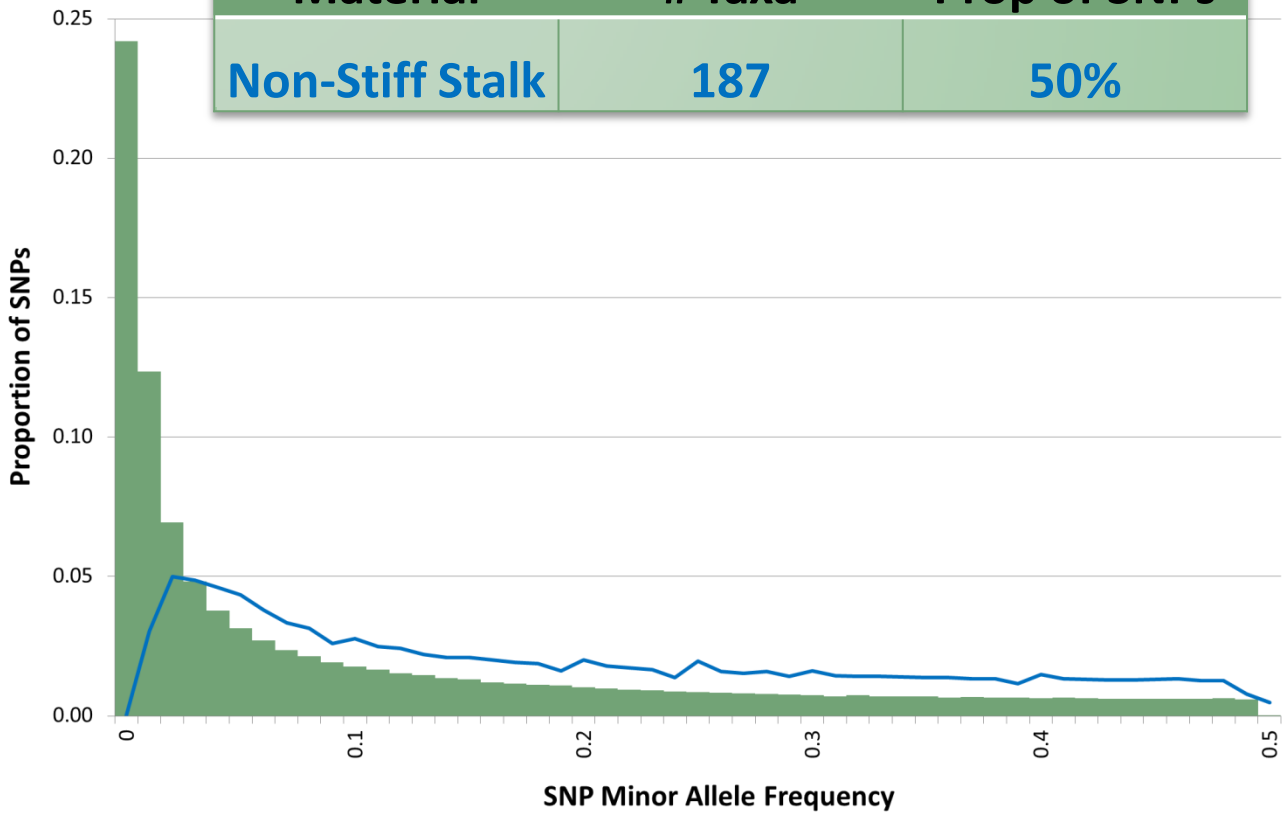
SNP distribution across 2,709 lines vs. 402 lines with tropical origin

How are alleles distributed?

Non-Stiff Stalk have a substantial bottleneck



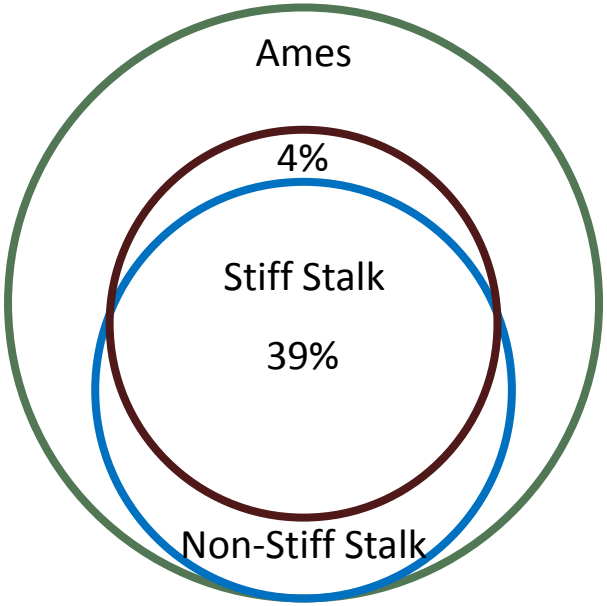
Material	# Taxa	Prop of SNPs
Non-Stiff Stalk	187	50%



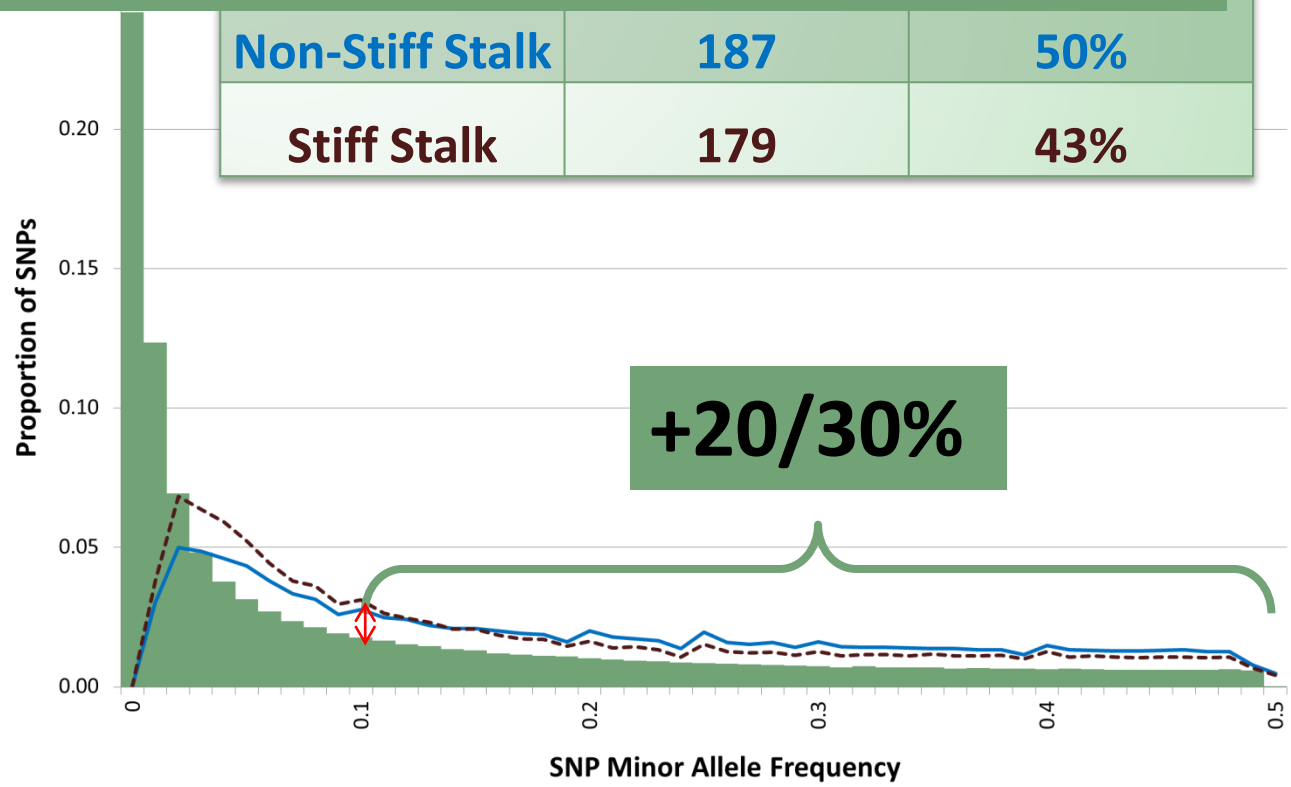
SNP distribution across 2,709 lines vs. 187 lines classified as Non-Stiff Stalk

How are alleles distributed?

Modern breeding has focused on working with common alleles



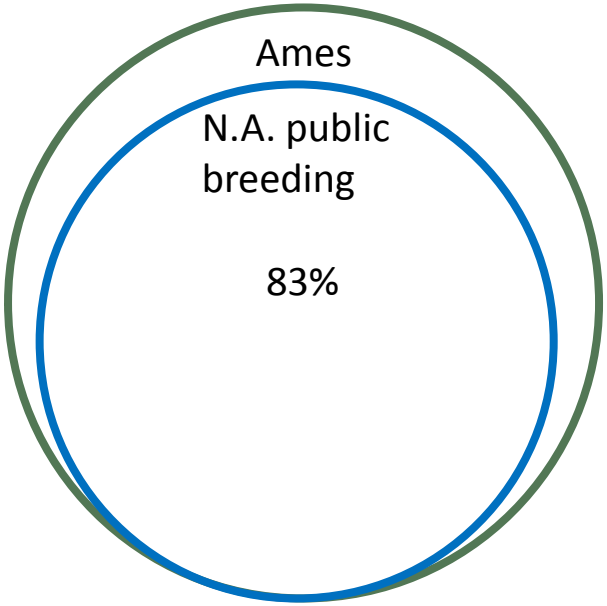
Non-Stiff Stalk	187	50%
Stiff Stalk	179	43%



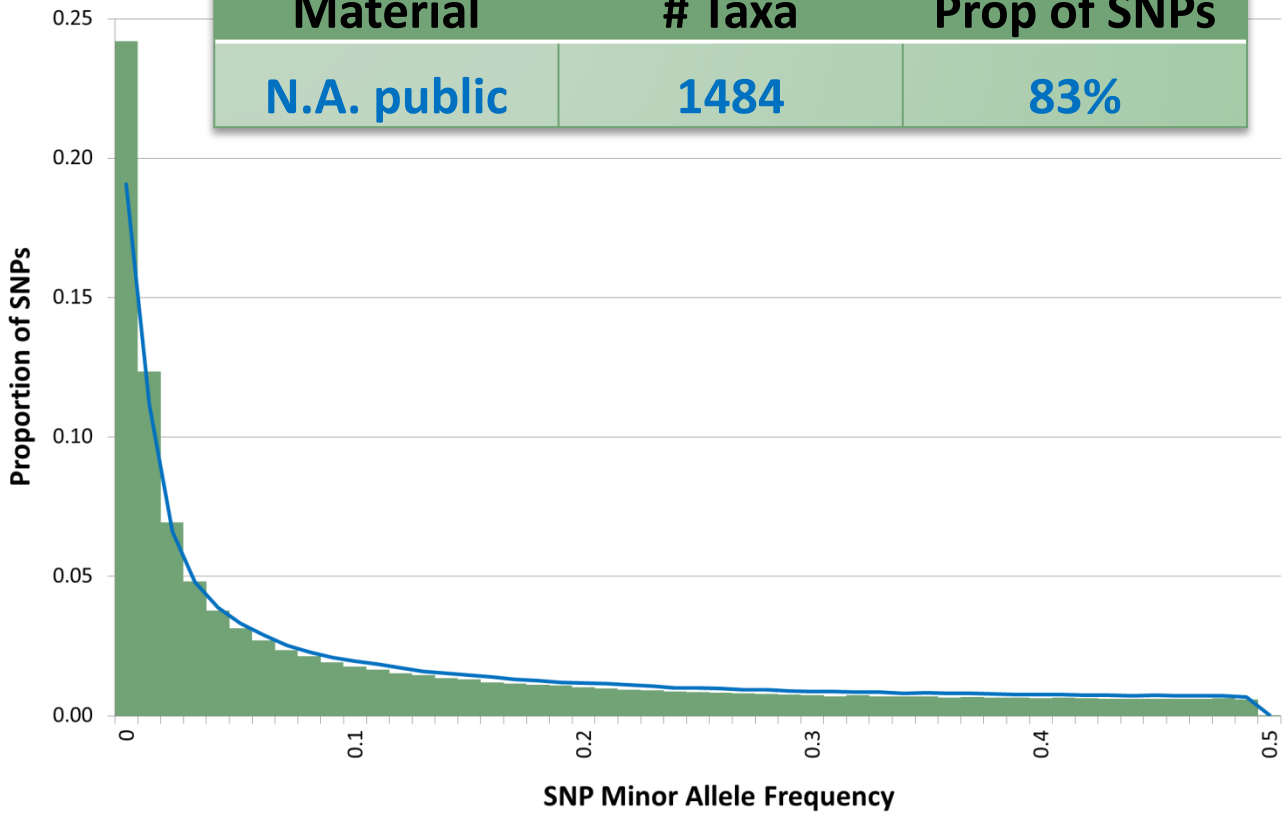
SNP distribution across 2,709 lines vs. 187 lines classified as Non-Stiff Stalk and 179 classified as Stiff Stalk

How are alleles distributed?

USA and Canada public breeding programs have done a good job adding diversity to those groups



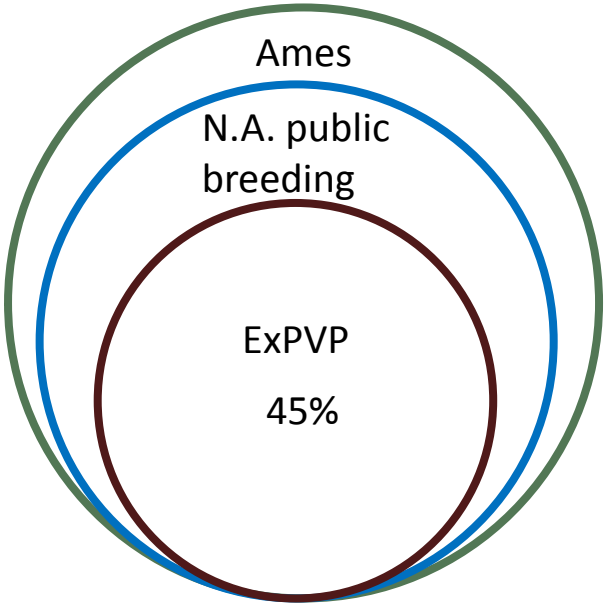
Material	# Taxa	Prop of SNPs
N.A. public	1484	83%



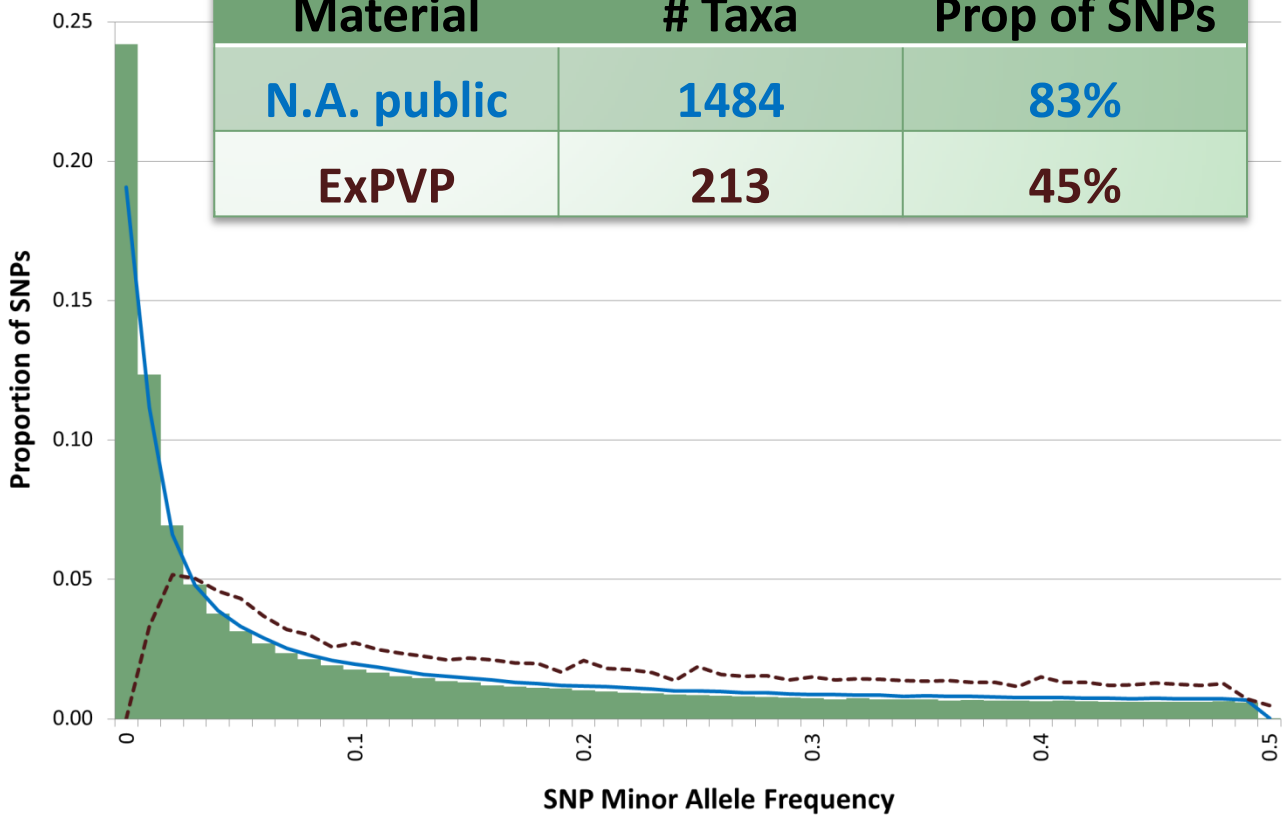
SNP distribution across 2,709 lines vs. 1484 USA and Canadian lines from public breeding programs (GEM, NC breeding program, etc.)

How are alleles distributed?

But a modest amount of available diversity has been commercially exploited



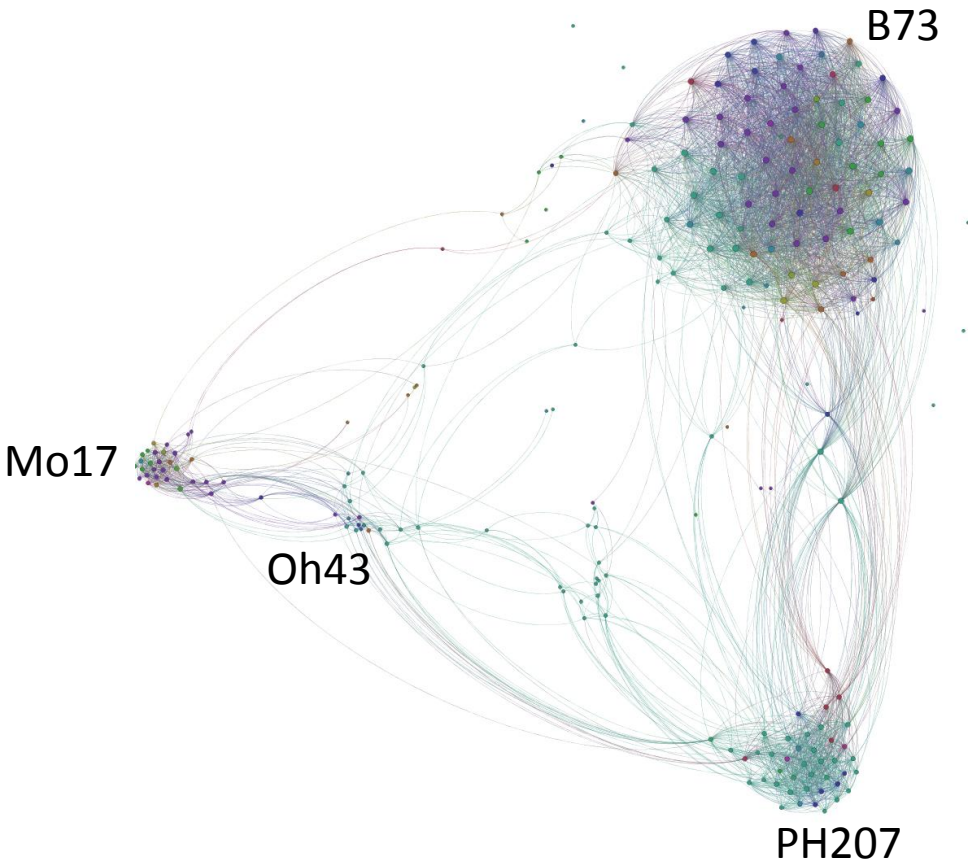
Material	# Taxa	Prop of SNPs
N.A. public	1484	83%
ExPVP	213	45%



SNP distribution across 2,709 lines vs. 1484 N.A lines from public breeding programs and 213 ExPVP

How are alleles distributed?

Breeding has pushed alleles apart



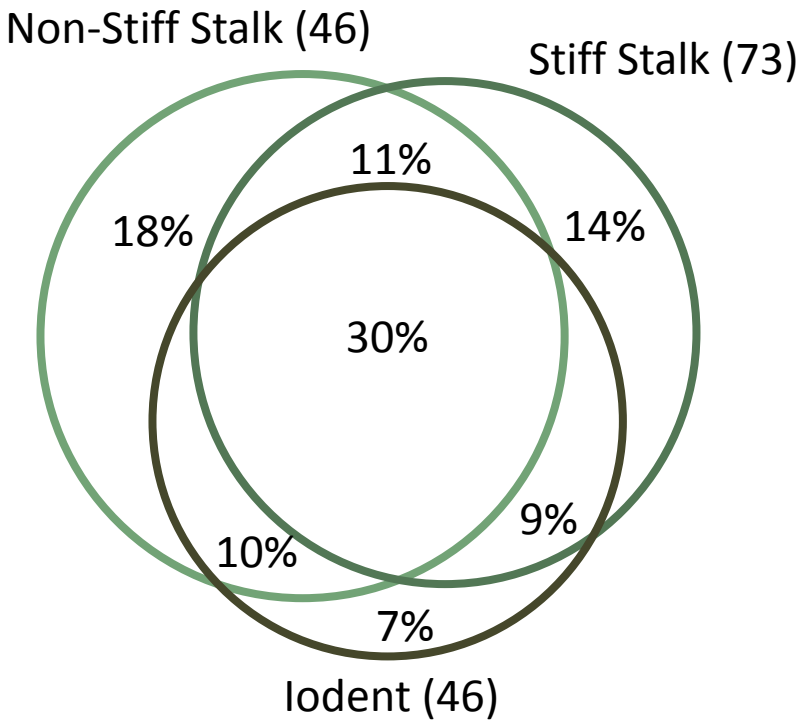
ExPVPs from different companies form clusters around 4 inbreds

Network relations for inbreds sharing more than 90 % of SNPs

How are alleles distributed?

Breeding has pushed alleles apart

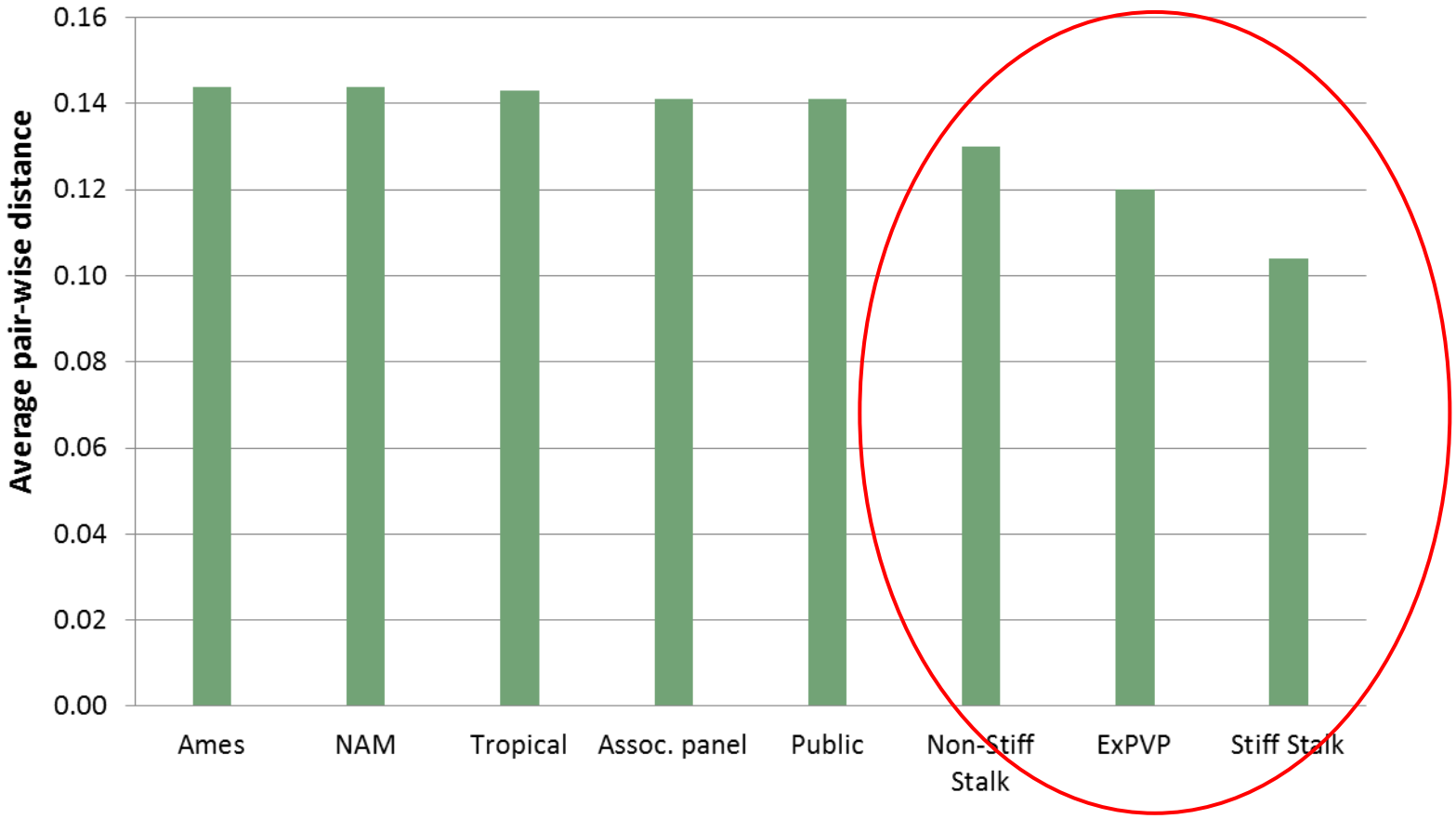
70% of alleles are not shared between groups



Allele distribution between 3 main heterotic pools

How are alleles distributed?

????????



Average genetic distances within groups of germplasm

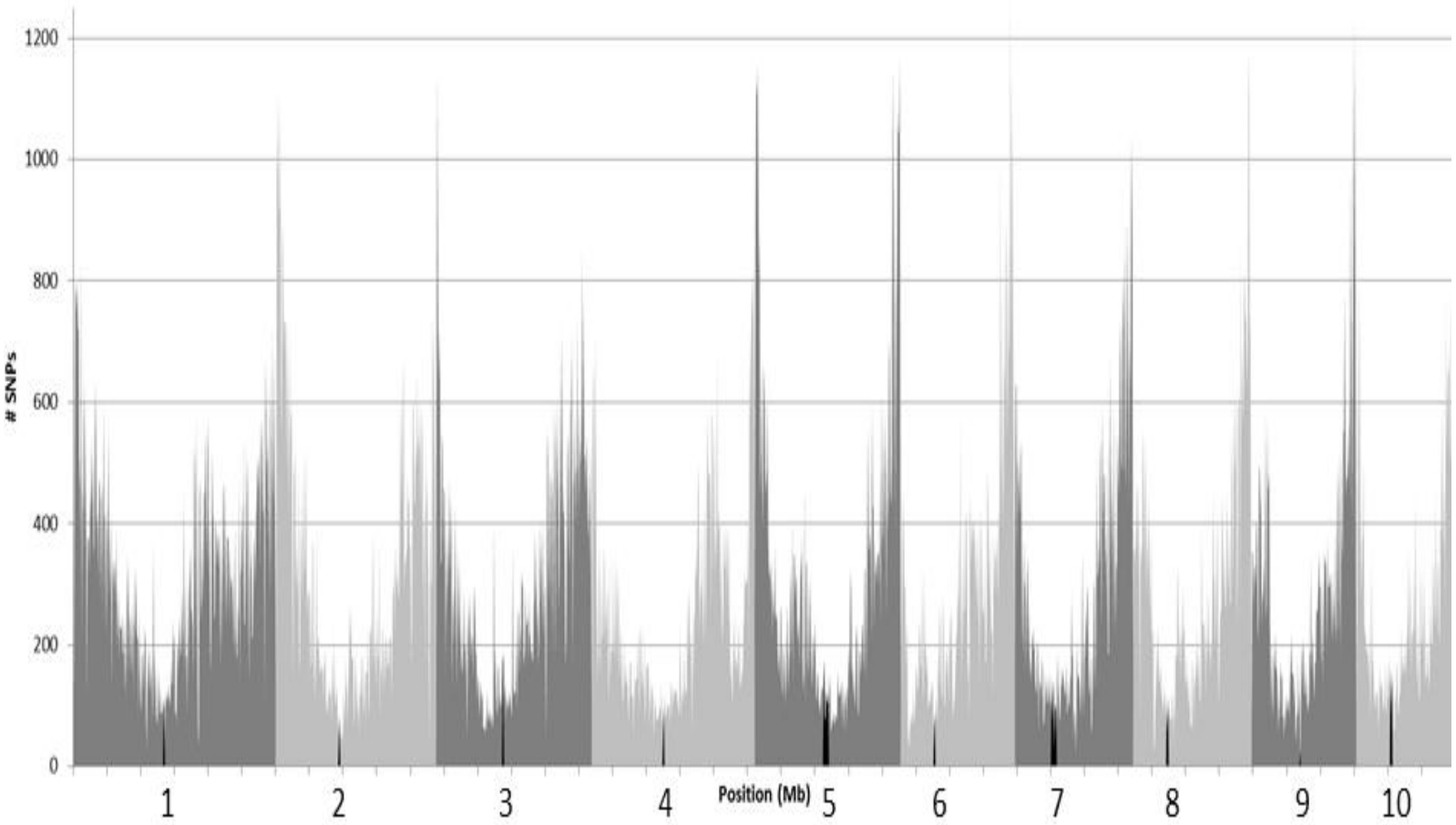


Fig.1. Distribution of number of SNPs found in 1 Mb windows across the 10 maize chromosomes. Centromere positions are showed in black.

Conclusions

- Ames panel captures an important slice of rare alleles
- Temperate materials, specially ExPVPs, only capture a modest amount of the available diversity. Tropical inbreds, landraces and Teosinte can be added to increase diversity
- GBS data can assist with curational management

Goals: Application of Technology to Identification and Capture of Valuable Allelic Diversity

Phenotyping

Genotyping

- GBS
- Resequencing

Association analysis

GWAS

Bioinformatics

Leveraging information across platforms

The germplasm: Ames inbred panel

TOTAL: 2,812 unique taxa (2,711 from the USDA collection in Ames, IA)

Includes different breeding materials:

- 213 ExPVP
- 212 GEM (Germplasm Enhancement of Maize)
- 179 classified as Stiff Stalk
- 187 classified as Non-Stiff Stalk
- 141 sweet corn
- 80 popcorn
- 402 from tropical origin

Different breeding origins:

- 31 different countries (USA, Spain, Mexico, Thailand, China, Nigeria, etc.)
- 35 different US states
- Almost a century of breeding efforts

C. Romay, 2012

Maize Curator Develops Query Method for Studying Relationships between Inbred Lines

Cinta Romay provided a divergence matrix based on hundreds of thousands of SNP loci to Curator Mark Millard. The value is a % different SNPs based on good reads.

4476 rows by 4476 columns.

	A	EH	EI	EJ	EK	EL	EM	EN
1	Taxa	H49_Ames15929_03ncai01_SD	Ab28A_Ames18999_04ncai02_SD	Ab28A_Ames18999_07ncai01_SD	VaW6_Ames19000_07ncai01_SD	VaW6_Ames19000_04ncai02_SD	VaW6_Ames19000_11336	VaW6_Ames19000_10775
2	12E_12E_Edwards	0.134917494	0.139242293	0.140492007	0.145038742	0.143879286	0.146195418	0.144261778
3	3167B_3167B_Edwards	0.130183161	0.137636892	0.138651868	0.141425722	0.143182259	0.145864336	0.140936507
4	3316_282	0.143095105	0.145715499	0.145730685	0.145550476	0.143303166	0.143890841	0.143880475
5	3316_282	0.14189895	0.144608311	0.144948982	0.145577194	0.146858058	0.141297673	0.14489338
6	3811_282	0.137110723	0.088557535	0.088496925	0.141087038	0.140873478	0.143455085	0.139328277
7	3811_282	0.137184284	0.088583733	0.089397509	0.141804576	0.142356599	0.146376294	0.141611335
8	4226_282	0.138628819	0.142423394	0.142937476	0.146171892	0.14936459	0.147541552	0.147152628
9	4226_282	0.135987423	0.14129441	0.141716748	0.144144582	0.147113534	0.147815168	0.145391267
10	4722_282	0.150684533	0.150009198	0.1514589	0.152198162	0.154102592	0.157323288	0.152968411
11	4722_282	0.150486288	0.150879975	0.151058452	0.153650331	0.15599405	0.157095727	0.153473415
12	A131_A131_Holland	0.138924402	0.143309239	0.142978859	0.144694841	0.144546248	0.146857031	0.145412662
13	A131 (Holland)_A131(Holland)_10371	0.153139242	0.151500223	0.152800975	0.152115233	0.163785951	0.155966916	0.15661231
14	A188_282	0.148102064	0.150075273	0.150258229	0.15054522	0.14856424	0.155901093	0.150307447
15	A188_282	0.144355662	0.14783265	0.14847165	0.14930812	0.146638596	0.151687222	0.150037423
16	A214N_282	0.145635588	0.145908952	0.147399281	0.144469798	0.145114237	0.145068694	0.145651399
17	A214N_282	0.142552257	0.146307016	0.14721141	0.14453952	0.146256489	0.145743223	0.145044507
18	A239_282	0.139877604	0.142195992	0.14361971	0.148330853	0.154586835	0.147608979	0.151515152
19	A239_282	0.136064912	0.141430739	0.142938607	0.147248359	0.148780892	0.146527778	0.147956274
20	A272_282	0.157141685	0.156391743	0.156960311	0.158489569	0.153452241	0.157467158	0.157738774
21	A272_282	0.158558179	0.157057784	0.158980639	0.161662798	0.15857356	0.162906708	0.160945163
22	A3G-3-1-3_A3G-3-1-3_Edwards	0.154867984	0.150817854	0.152803819	0.155442187	0.154952485	0.158802798	0.157240119
23	A4415_282	0.147061947	0.148823865	0.149295151	0.148967048	0.149010675	0.151526959	0.148400671
24	A4415_282	0.144797826	0.146132116	0.147112396	0.147268633	0.149291796	0.152779482	0.149608235
25	A554_282	0.110015387	0.142587134	0.143254758	0.147172408	0.142545057	0.145375972	0.148329786
26	A554_282	0.10874335	0.140568114	0.142858497	0.145610581	0.145047995	0.143496565	0.144207924
27	A556_282	0.141492578	0.143771313	0.144324922	0.144835883	0.14648515	0.145883547	0.143132507
28	A556_282	0.140172984	0.142398418	0.142372389	0.14529602	0.146629011	0.145741525	0.143575175
29	A6_282	0.154034338	0.154267395	0.154600812	0.151167084	0.148540508	0.152353383	0.150482563
30	A6_282	0.153860583	0.151999884	0.153973487	0.150920763	0.148704333	0.152975786	0.153098209
31	A619_282	0.142320293	0.146791286	0.145488618	0.12594774	0.12454917	0.128917321	0.124749477
32	A619_282	0.142035054	0.143809925	0.144090331	0.12508651	0.124500366	0.127861505	0.124905141
33	A632_282	0.134228442	0.136580296	0.139273506	0.140617612	0.140311463	0.142294338	0.140811633
34	A632_282	0.131776065	0.139029087	0.139176427	0.140031134	0.140887141	0.142917402	0.143565087
35	A634_282	0.135234858	0.138858424	0.140143302	0.139841689	0.139719748	0.142297185	0.141938904
36	A634_282	0.133007191	0.136824242	0.138239817	0.138617768	0.13839644	0.141449372	0.139541539

Matrix_Ames_120110 | Example

That data was just too hard to handle so he made an entry list table and then created a mapping table that had this value in it. Those two tables looked like this. The germplasm1_id and germplasm2_id have the snp_entry_id in them.

Microsoft SQL Server Management Studio

File Edit View Query Project Debug Tools Window Community Help

New Query | master | Execute

Object Explorer

Connect

- dbo.load0200u
- dbo.method
- dbo.method_map
- dbo.name_group
- dbo.order_request
- dbo.order_request_action
- dbo.order_request_attach
- dbo.order_request_item
- dbo.region
- dbo.site
- dbo.snp_entry
  - Columns
    - snp_entry_id (PK, int, not null)
    - taxa (nvarchar(255), null)
    - snp_entry_id_saved (int, null)
    - created_date (datetime2(7), not null)
    - created_by (int, not null)
    - modified_date (datetime2(7), null)
    - modified_by (int, null)
    - owned_date (datetime2(7), not null)
    - owned_by (int, not null)
  - Keys
  - Constraints
  - Triggers
  - Indexes
  - Statistics
- dbo.snp_entry_map
- dbo.sys_database
- dbo.sys_database_migration
- dbo.sys_database_migration_lang
- dbo.sys_datatrigger
- dbo.sys_datatrigger_lang
- dbo.sys_dataview
- dbo.sys_dataview_field
- dbo.sys_dataview_field_lang
- dbo.sys_dataview_lang
- dbo.sys_dataview_param
- dbo.sys_dataview_sql

Toolbox

General

SQLQuery1.sql -...\mjmillar (53))

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP 1000 [snp_entry_id]
      ,[taxa]
      ,[snp_entry_id_saved]
      ,[created_date]
      ,[created_by]
      ,[modified_date]
      ,[modified_by]
      ,[owned_date]
      ,[owned_by]
  FROM [gringlobal].[dbo].[snp_entry]
```

Results | Messages

	snp_entry_id	taxa	snp_entry_id_saved	created_date	created_by	modified_date	modified_by	owned_date	owned
73	73	GE129_Ames10256_12833	73	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
74	74	GE281_Ames10257_90ncab01_SD	74	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
75	75	GE293_Ames10258_90ncab01_SD	75	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
76	76	GE311_Ames10259_07ncei01_SD	76	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
77	77	GE317_Ames10260_96ncab01_SD	77	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
78	78	GE335_Ames10261_90ncab01_SD	78	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
79	79	GT150_Ames10262_90ncab01_SD	79	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
80	80	L 90_Ames10263_90ncab01_SD	80	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
81	81	L 605_Ames10264_90ncab01_SD	81	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
82	82	L 678_Ames10265_90ncab01_SD	82	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
83	83	L 690_Ames10266_12872	83	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
84	84	SC15_Ames10267_90ncab01_SD	84	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
85	85	SC16_Ames10268_12894	85	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
86	86	SC44_Ames10269_91ncai01_SD	86	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
87	87	SC84E_Ames10271_91ncai01_SD	87	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
88	88	SC90_Ames10272_91ncai01_SD	88	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
89	89	SC102_Ames10273_96ncab01_SD	89	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
90	90	SC144_Ames10274_91ncai01_SD	90	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
91	91	SC144_Ames10274_12895	91	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
92	92	SC225_Ames10276_08ncai01_SD	92	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1
93	93	SC235_Ames10277_96ncab01_SD	93	2012-06-09 11:16:46.6493639	1	NULL	NULL	2012-06-09 11:16:46.6493639	1

Microsoft SQL Server Management Studio

File Edit View Query Project Debug Tools Window Community Help

New Query | master | Execute

Object Explorer

Connect

- owned_date (datetime2(7), not null)
- owned_by (int, not null)
- Keys
- Constraints
- Triggers
- Indexes
- Statistics
- dbo.snp_entry_map
  - Columns
    - snp_entry_map_id (PK, int, not null)
    - germplasm1_id (int, null)
    - germplasm2_id (int, null)
    - divergence (float, null)
    - similarity (float, null)
    - created_date (datetime2(7), not null)
    - created_by (int, not null)
    - modified_date (datetime2(7), null)
    - modified_by (int, null)
    - owned_date (datetime2(7), not null)
    - owned_by (int, not null)
  - Keys
  - Constraints
  - Triggers
  - Indexes
  - Statistics
- dbo.sys_database
- dbo.sys_database_migration
- dbo.sys_database_migration_lang
- dbo.sys_datatrigger
- dbo.sys_datatrigger_lang
- dbo.sys_dataview
- dbo.sys_dataview_field
- dbo.sys_dataview_field_lang
- dbo.sys_dataview_lang
- dbo.sys_dataview_param
- dbo.sys_dataview_sql
- dbo.sys_file
- dbo.sys_file_group

Toolbox

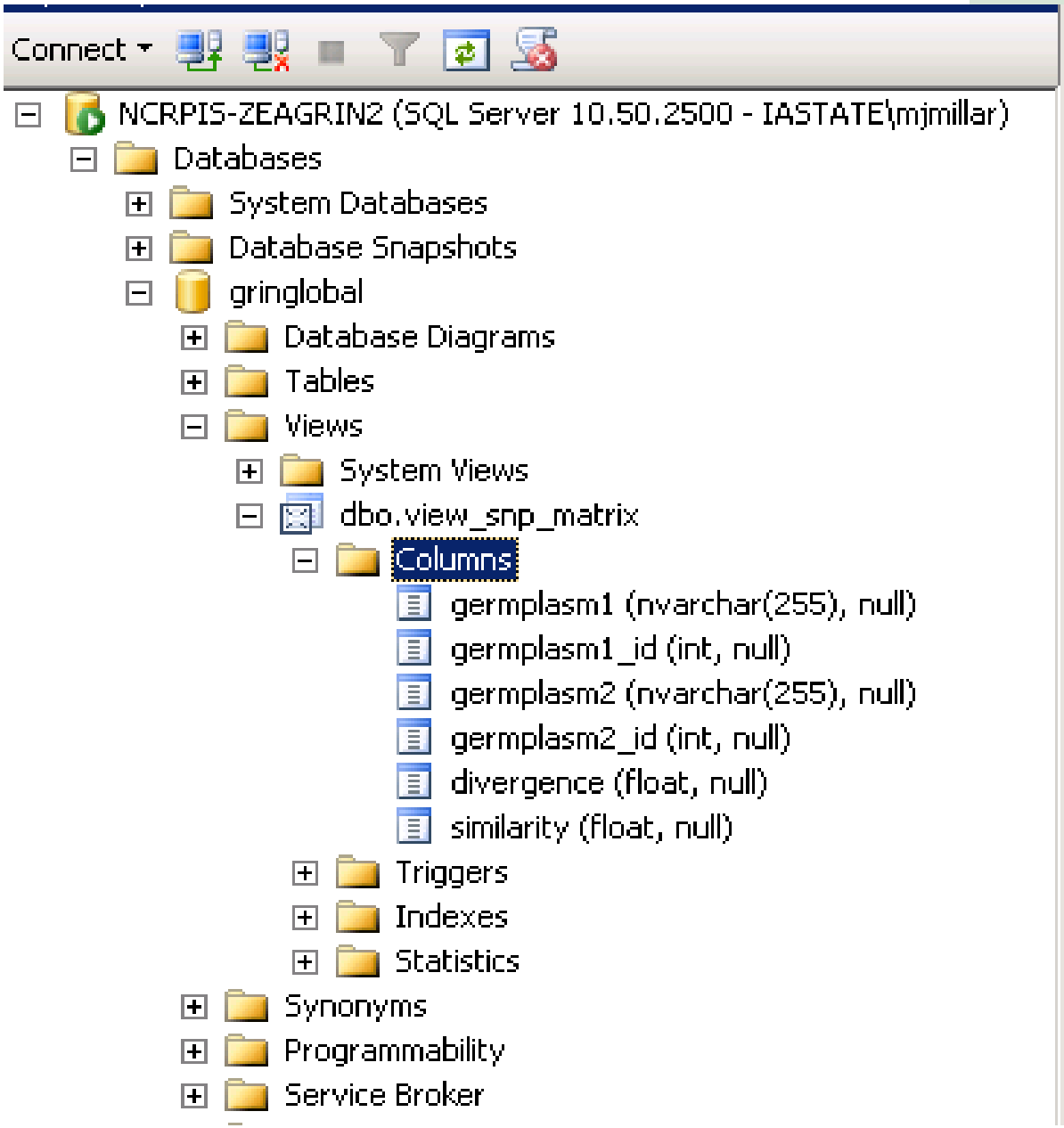
General

SQLQuery2.sql -...\mjmillar (51)) | SQLQuery1.sql -...\mjmillar (53))

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP 1000 [snp_entry_map_id]
      ,[germplasm1_id]
      ,[germplasm2_id]
      ,[divergence]
      ,[similarity]
      ,[created_date]
      ,[created_by]
      ,[modified_date]
      ,[modified_by]
      ,[owned_date]
      ,[owned_by]
  FROM [gringlobal].[dbo].[snp_entry_map]
```

Results | Messages

	snp_entry_map_id	germplasm1_id	germplasm2_id	divergence	similarity	created_date	created_by	modified_date	modified_by	owned_date
1	36703202	1	1	0	1	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
2	36703203	2	1	0.132309749722481	0.867690250277519	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
3	36703204	3	1	0.141625478863716	0.858374521136284	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
4	36703205	4	1	0.142576202750206	0.857423797249794	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
5	36703206	5	1	0.135260254144669	0.864739745855331	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
6	36703207	6	1	0.1359493881464	0.8640506118536	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
7	36703208	7	1	0.13973993062973	0.86026006937027	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
8	36703209	8	1	0.137673273682594	0.862326726317406	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
9	36703210	9	1	0.148422554135323	0.851577445864677	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
10	36703211	10	1	0.149229913949966	0.850770086050034	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
11	36703212	11	1	0.13839328289032	0.86160671710968	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
12	36703213	12	1	0.144048735499382	0.855951264500618	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
13	36703214	13	1	0.148141279816628	0.851858720183373	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
14	36703215	14	1	0.146216258406639	0.853783741593361	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
15	36703216	15	1	0.141457542777061	0.858542457222939	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
16	36703217	16	1	0.13909724354744	0.86090275645256	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
17	36703218	17	1	0.141469404101372	0.858530595898628	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
18	36703219	18	1	0.140780001878738	0.859219998121262	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1
19	36703220	19	1	0.153298482298851	0.846701517701149	2012-06-13 14:43:25.9404160	1	NULL	NULL	2012-06-13 1



He then created a SQL Server view that looks like this:

Then he created an ODBC (open database connectivity) connection to the database with Excel. The top allows entry of parameters that the spreadsheet below then presents.

Germplasm 1 Like		%27126%	
Germplasm 2 Like		%%	
Divergence >=		0	
Similarity <=		1	

germplasm1	germplasm1_id	germplasm2	germplasm2_id	divergence	similarity
Ki43_Ames27126_12393	825	Ki43_282	2007	0.0064948	0.993505
Ki43_Ames27126_12393	825	Ki43_Ames27126_04ncai01_SD	824	0.0072390	0.992760
Ki43_Ames27126_04ncai01	824	Ki43_Ames27126_12393	825	0.0072390	0.992760
Ki43_Ames27126_04ncai01	824	Ki43_282	2007	0.0091088	0.990891
Ki43_Ames27126_04ncai01	824	Ki43_282	2008	0.0102069	0.989793
Ki43_Ames27126_12393	825	Ki2021_282	1996	0.1192408	0.880759
Ki43_Ames27126_04ncai01	824	Ki2021_282	1996	0.1194020	0.880597
Ki43_Ames27126_04ncai01	824	CML481_CML481_Krakowksy	1608	0.1205745	0.879425

Understanding how Useful Diversity was Captured

Wf9	Indiana Station strain of Reid Yellow Dent
W64A	Wf9 x C.I.187-2
R222	(W64A x Ht1B) W64A BC3
R222A	(W64A x Ht1B) W64A BC3
SD41	SDp309 x W64A
SD60	White conversion of W64A using SD316W
AusTRCF 306335	{W64Abm3 x Gn32}-1-1-1-1-1
C.I. 187-2	Krug
K63	Pride of Saline
SDp309	SDp236 X K63
SD316W	SDp236 x K63
SDp236M	Fulton Yellow Dent

Millards Query to ZEAGRIN2.xlsx

	A	B	C	D	E	F	G
1	Germplasm1 Like		%W64A_PI587152_94ncai01_SD%				Statistics to watch!
2	Germplasm2 Like		%%				
3	Divergence >=		0				B73 and Mo17 are 86% similar
4	Similarity <=		1				B73 and Till, Teosinte Lines is 79.7% similar
5							DNA samples from one lot B73_Flint-Garcia go from 99.79% to 98.13% similar with mean 99.56% standard dev of .36%
6							
10	germplasm1	germplasm1_id	germplasm2	germplasm2_id	divergence	similarity	
11	W64A_PI587152_94ncai01_SD	3518	W64A_282	4468	0.00387828	0.99612172	
12	W64A_PI587152_94ncai01_SD	3518	W64A_282	4469	0.005062772	0.994937228	
13	W64A_PI587152_94ncai01_SD	3518	W64A_NSL30058_05ncai01_SD	2312	0.013472688	0.986527312	
14	W64A_PI587152_94ncai01_SD	3518	W64A ae/wx_Ames26768_03ncai01_SD	642	0.018175215	0.981824785	
15	W64A_PI587152_94ncai01_SD	3518	W64A T_Ames19292_91ncao01_SD	179	0.018325064	0.981674936	
16	W64A_PI587152_94ncai01_SD	3518	W64A ae_Ames26767_03ncai01_SD	641	0.018714802	0.981285198	
17	W64A_PI587152_94ncai01_SD	3518	R222A_NSL29898_07ncai01_SD	2301	0.026120329	0.973879671	
18	W64A_PI587152_94ncai01_SD	3518	R222_NSL29319_95ncai01_SD	2300	0.030645607	0.969354393	
19	W64A_PI587152_94ncai01_SD	3518	SD60_PI538236_99ncai01_SD	3102	0.031268805	0.968731195	
20	W64A_PI587152_94ncai01_SD	3518	SD41_PI606769_91ncai01_SD	3929	0.045658555	0.954341445	
21	W64A_PI587152_94ncai01_SD	3518	SD41_PI606769_91ncai01_SD	3931	0.045880444	0.954119556	
22	W64A_PI587152_94ncai01_SD	3518	SD41_PI606769_91ncai01_SD	3933	0.046366338	0.953633662	
23	W64A_PI587152_94ncai01_SD	3518	SD41_PI606769_91ncai01_SD	3930	0.04645123	0.95354877	
24	W64A_PI587152_94ncai01_SD	3518	SD41_PI606769_91ncai01_SD	3932	0.04657514	0.95342486	
25	W64A_PI587152_94ncai01_SD	3518	SD41_PI606769_91ncai01_SD	3928	0.048183542	0.951816458	
26	W64A_PI587152_94ncai01_SD	3518	AusTRCF 306335_NSL438021_09ncai01_SD	2507	0.059017844	0.940982156	
27	W64A_PI587152_94ncai01_SD	3518	Wf9_282	4473	0.060593382	0.939406618	
28	W64A_PI587152_94ncai01_SD	3518	Wf9_Ames19293_07ncai01_SD	180	0.061894748	0.938105252	
29	W64A_PI587152_94ncai01_SD	3518	Wf9_282	4474	0.062096361	0.937903639	
30	W64A_PI587152_94ncai01_SD	3518	WXWF9_Ames26758_03ncai01_SD	629	0.062646031	0.937353969	
31	W64A_PI587152_94ncai01_SD	3518	WXWF9_Ames26758_03ncai01_SD	631	0.063081875	0.936918125	
32	W64A_PI587152_94ncai01_SD	3518	WXWF9_Ames26758_03ncai01_SD	632	0.063215576	0.936784424	
33	W64A_PI587152_94ncai01_SD	3518	WXWF9_Ames26758_03ncai01_SD	628	0.063674532	0.936325468	
34	W64A_PI587152_94ncai01_SD	3518	WXWF9_Ames26758_03ncai01_SD	627	0.064008735	0.935991265	
35	W64A_PI587152_94ncai01_SD	3518	Wf9 T_Ames19294_91ncao01_SD	181	0.064626671	0.935373329	
36	W64A_PI587152_94ncai01_SD	3518	WXWF9_Ames26758_03ncai01_SD	630	0.064948142	0.935051858	
37	W64A_PI587152_94ncai01_SD	3518	R909_NSL30909_07ncai01_SD	2379	0.068656638	0.931343362	
38	W64A_PI587152_94ncai01_SD	3518	R909TMS_NSL29286_01ncfo01_SD	2290	0.069564812	0.930435188	
39	W64A_PI587152_94ncai01_SD	3518	A630_Ames23466_98ncai01_SD	435	0.070495568	0.929504432	

Sheet1 Sheet2 Sheet3

Gardner B73 Similarity.xlsx - Microsoft Excel

	A	B	C	D	E	F
1	Germplasm1 Like		%B73_PI550473_08ncai02_SD%			
2	Germplasm2 Like		%%			
3	Divergence >=		0			
4	Similarity <=		1			
5						
6						
7	germplasm1	germplasm1_id	germplasm2	germplasm2_id	divergence	similarity
8	B73_PI550473_08ncai02_SD	3278	B73_282	1205	0.002063222	0.99793678
9	B73_PI550473_08ncai02_SD	3278	B73_PI550473_10099	3279	0.002585488	0.99741451
10	B73_PI550473_08ncai02_SD	3278	B73_282	1206	0.002863302	0.9971367
11	B73_PI550473_08ncai02_SD	3278	GEMS-0101_GEMS-0101_12214	1894	0.003039401	0.9969606
12	B73_PI550473_08ncai02_SD	3278	B73_B73_Flint-Garcia	1224	0.003216751	0.99678325
13	B73_PI550473_08ncai02_SD	3278	F42_PI601026_08ncai01_SD	3730	0.003715402	0.9962846
14	B73_PI550473_08ncai02_SD	3278	B73Htrhm_Ames27065_03ncai01_SD	739	0.003834897	0.9961651
15	B73_PI550473_08ncai02_SD	3278	B73_PI550473_08ncai01_SD	3277	0.003955584	0.99604442
16	B73_PI550473_08ncai02_SD	3278	DJ7_PI601191_04ncai01_SD	3741	0.005011542	0.99498846
17	B73_PI550473_08ncai02_SD	3278	NC372_Ames27182_04ncai01_SD	918	0.005892336	0.99410766
4431	B73_PI550473_08ncai02_SD	3278	TIL08_Teo	4346	0.192192405	0.80780759
4432	B73_PI550473_08ncai02_SD	3278	TIL15_Teo	4374	0.192248732	0.80775127
4433	B73_PI550473_08ncai02_SD	3278	TIL06_Teo	4337	0.192329466	0.80767053
4434	B73_PI550473_08ncai02_SD	3278	TIL16_Teo	4382	0.192599028	0.80740097
4435	B73_PI550473_08ncai02_SD	3278	TIL16_Teo	4380	0.19273147	0.80726853
4436	B73_PI550473_08ncai02_SD	3278	TIL08_Teo	4348	0.192810059	0.80718994
4437	B73_PI550473_08ncai02_SD	3278	TIL08_Teo	4344	0.192932382	0.80706762
4438	B73_PI550473_08ncai02_SD	3278	TIL15_Teo	4376	0.19320403	0.80679597
4439	B73_PI550473_08ncai02_SD	3278	TIL08_Teo	4345	0.193233341	0.80676666
4440	B73_PI550473_08ncai02_SD	3278	TIL15_Teo	4378	0.193510458	0.80648954
4441						

Schema considerations to handle SNP data and curatorial/public use of this data:

- add a second inventory_id link in the crop_trait_observation table in order to make the crop_trait_observation table sort of a mapping table.

- if GRIN could handle all the datapoints on the SNPs and had enough computing power then the similarity could be calculated on the fly. This table is pretty snappy to query. Computing power required to do it on the fly?

The 1,000 year plan would be to have this capability to get similarities across all traits, phenotypic, genotypic, geographic, etc.

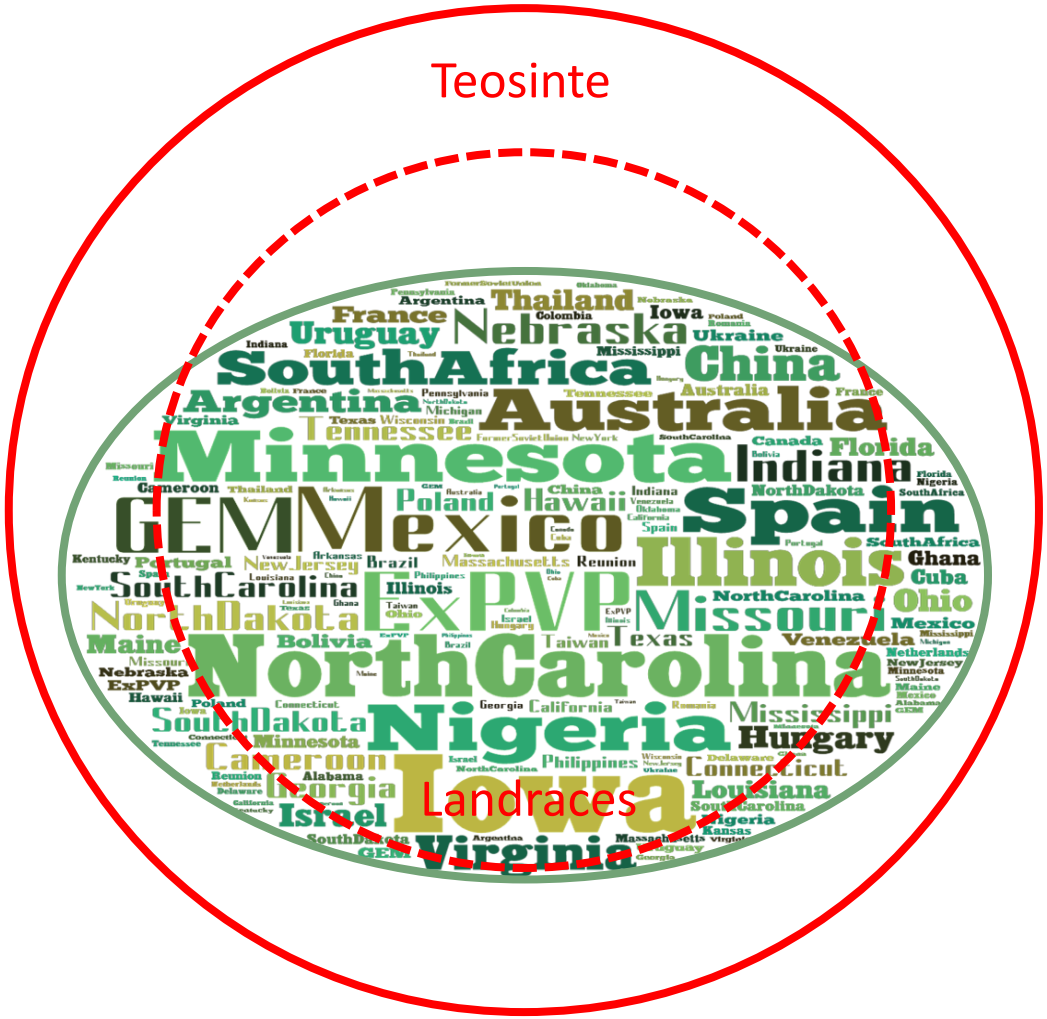
John Doebley lab website:

[a] We obtained three separate and conflicting pedigrees for M14 (a.k.a. Gn2; Don Duvick and Stan Jensen from Pioneer both indicated that Gn2 was derived from M14). (1) According to Bob Lambert at the University of Illinois (and Clarion Henderson of Illinois Foundation Seeds and Gerdes et.al., 1993), the pedigree is BR10 X R8, where BR10 is from Funk's Yellow Dent and R8 is from Texas Surecropper. (2) According to Tom Gocken at DeKalb, the pedigree is Lancaster X A, where A is a line from Funk's Yellow Dent. (3) According to records at Pioneer Hi-Bred, the pedigree is Lancaster X B21, where B21 is Fe2 X Steen 665. Ralph Hart of Moews Hybrids (Ben Moews developed M14 in the early 1930's) says that their existing records do not go back past 1966. Pioneer acquired Gn2 from Charlie Gunn of DeKalb Agricultural Research in 1935. DeKalb's records of it go back to 1932 (Tom Gocken, personal communication).

Conclusions

- The germplasm conserved at the PI Station in Ames combined with GBS information is a great resource to study the genetic architecture of quantitative traits as well as identify unique/similar accessions and resolve curatorial issues
- GBS data can be used to choose inbreds for different projects and help with the development of future breeding programs

For the future there is a huge world of extensive diversity to explore



When and where?

To obtain maize and/or wild *Zea* germplasm, contact:

Mark Millard (Maize Curator)
North Central Regional Plant Introduction Station (NC7) USDA/ARS & Iowa State University
State and Mortensen Rd.
Ames, IA 50011
E-Mail: Mark.Millard@ARS.USDA.GOV

GBS data will be publicly available following publication of results. Visit: http://www.panzea.org

Acknowledgements

- Buckler Lab:
  - Ed Buckler, Maria Cinta Romay
  - Bioinformatics (J. Harriman, J. Glaubitz, T. Casstevens, Y. Zhang, D. Kroon)
  - Statistics (Z. Zhang , A. Lipka, J. Peiffer); Field (N. Lepak)
- IGD (Institute for Genomic Diversity, Cornell University)
  - GBS team (S. Mitchell, R. Elshire, C .Acharya, W. Zhu, L. Blanchard)
- CBSU (Computational Biology Service Unit, Cornell University)
  - Q. Sun, R. Bukowski
- Flint-Garcia Lab
- Holland Lab
- McMullen Lab
- NCRPIS:
  - Mark Millard
  - Mike Blanco
- Major Goodman, Arnel Hallauer, Tom Hoegemeyer