

Next generation genotyping to enable use of maize genetic resources

Sarah Hearne,
Juan Burgueño,
J. Andrés Christen,
Jorge Franco,
Andrzej Kilian,
Ky L. Mathews,
Peter Wenzl

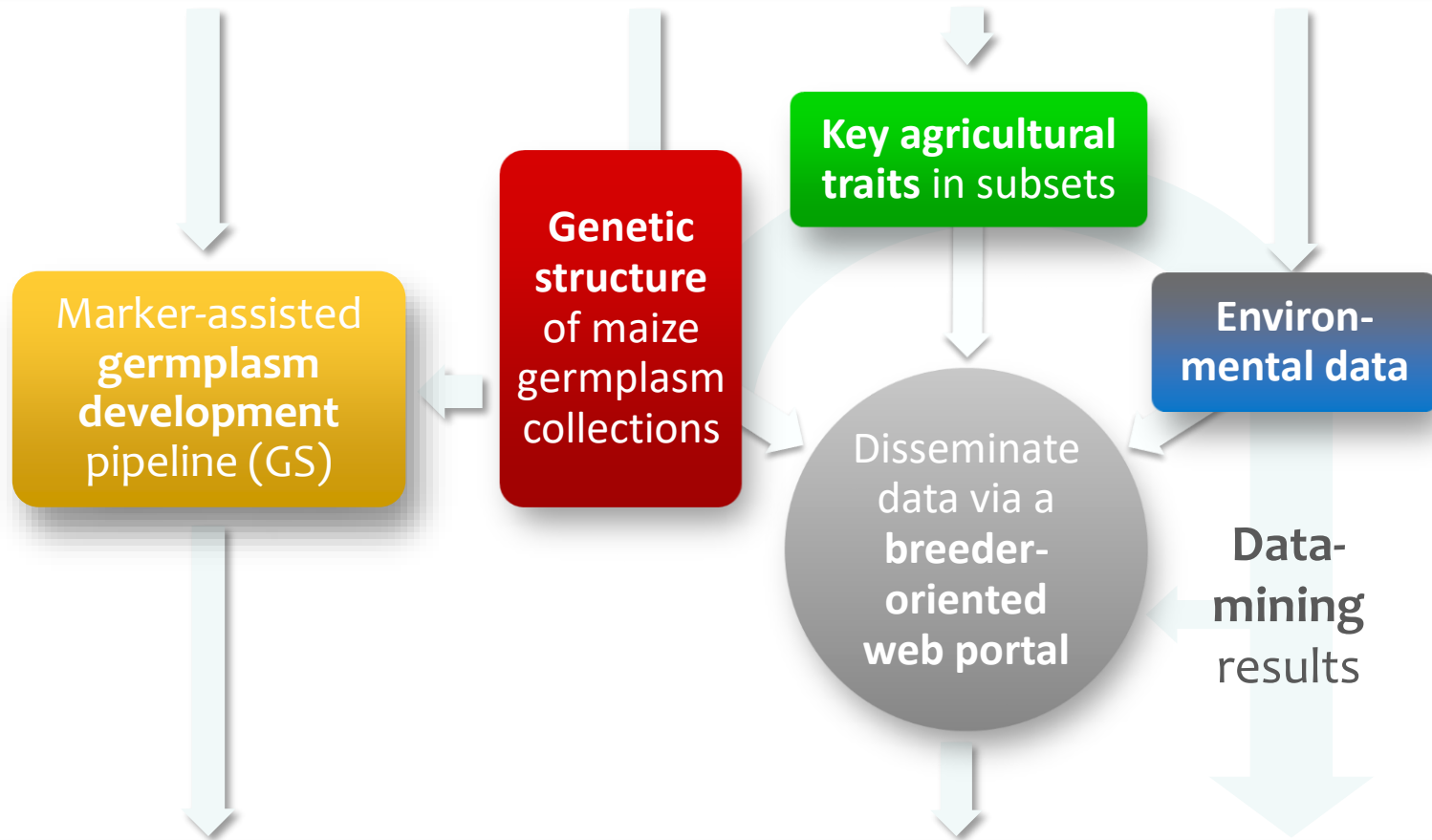


Genetic resources in genebanks



> 250,000 ancestral genotypes
in 45,000 populations
(accessions) at CIMMYT and
Mexican partner organizations.
Breeding lines, wild relatives.
(INIFAP, UAAAN, UACH, UdG)

Maize genetic resources

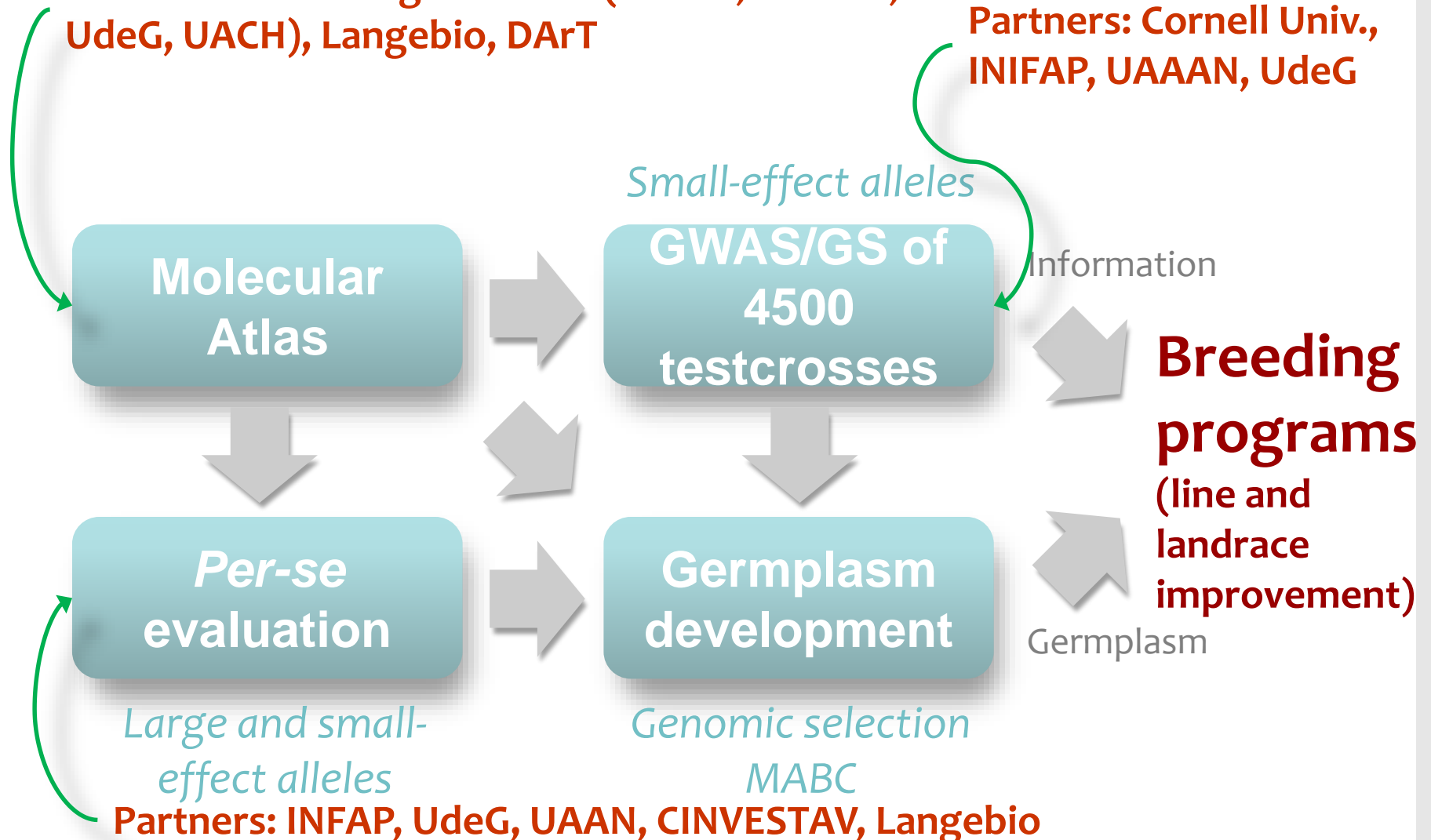


Global maize **breeding programs**

Maize strategy

Partners: Mexican genebanks (INIFAP, UAAAN, UdeG, UACH), Langebio, DArT

Partners: Cornell Univ., INIFAP, UAAAN, UdeG



Maize molecular atlas

Objective

- ▶ Understand and describe molecular genetic diversity of maize to aid germplasm selection and identify genetic imprints resulting from environmental and human influence

Outputs

- ▶ Fingerprints- genotypic profiles of maize landraces, elite materials and wild relatives available in a query-able portal with associated passport and characterization data where available

Technology

- SNP genotyping achieved through singleplex or low multiplex PCR based assays (TaqMan, KASP), pre-designed arrays like Illumina golden gate and Infinium, and thirdly through sequencing.
- Genotyping by sequencing (GBS) - scalable marker discovery and genotyping through next generation sequencing technologies. GBS - \$, tech advancement, repeatability (e.g. v exome capture).

Sampling

- Within accession variation in maize is large. Within accessions variance component $>$ between accession. How much variability do we want to capture?
- Resources available - \$, time, human and infrastructure capacity, technological capacity
- What questions do we want to ask of the data?
 - ▶ ...

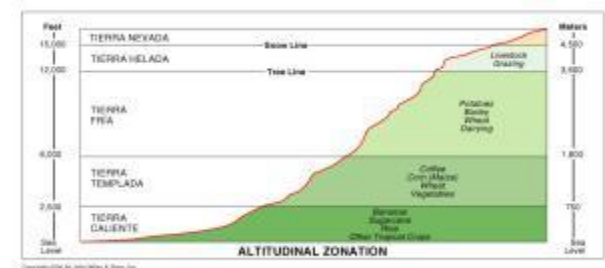
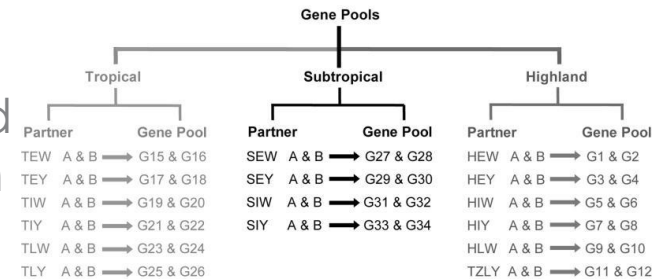
Potential questions

Some examples

- Germplasm bank curator
 - ▶ Race evaluation; “I have an accessions classified as pepitilla. Does it cluster with other accessions classified as pepitilla?”
- Breeder
 - ▶ Germplasm selection; “I want to expand my subtropical breeding pool to bring in some novel elite materials, which CMLs or other publically available materials could I use?”
- Evolutionary geneticists
 - ▶ Selection footprints “How does the frequency of allele/haplotype X change as altitude increases?”



Heterotic Partners of CIMMYT gene pools



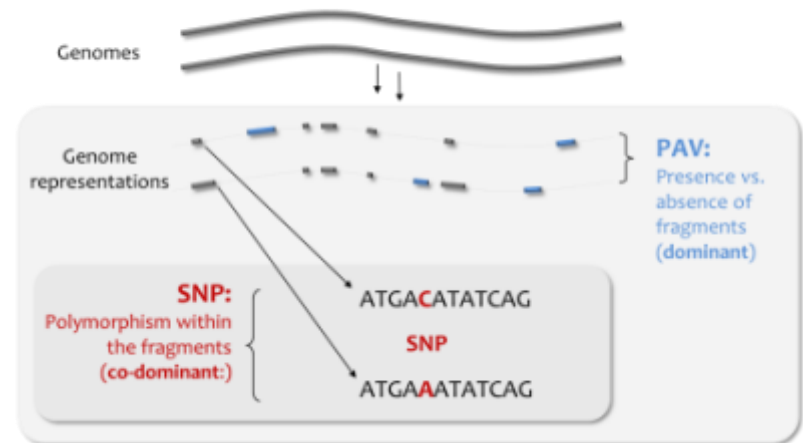
Maize genotyping criteria

- To avoid ascertainment bias due to sampling in heterogenous landraces we need to genotype multiple individuals- SSRs used 15 to 30 individuals
- To get an accurate representation of landraces we need to be able to score heterozygotes
- Reduce the impact of ascertainment bias due to SNP calling- reference inferred bias
- We need to do this in a resource effective manner

- IGD Cornell and DArT.
- IGD- Uses frequently cutting restriction enzyme (RE) to generate many fragments of genomic DNA which are then sequenced. More than 850,000 SNP are identified per sample in **homozygous** state.
- B73 reference
- Fantastic for bi-parental breeding applications and line evaluations. Imputation improves SNP density to between 1.5 to 2M SNP per sample and can be used to infer heterozygotes in structured populations
- Used for a large GWAS/GS study in SeeD
- Doesn't fit our criteria for large landrace application

- DArT GBS uses different RE systems than Cornell; two enzymes employed- one (PstI) has a partly overlapping recognition sequence with the RE employed by IGD
- Creates smaller number of fragments and therefore smaller number of SNP- C. 75,000 in total
- Can score heterozygotes in many loci as multiple copies of each fragment are sequenced (2-2.5M fragments sequenced per sample)
- PAV also generated
- Meets more specific needs

Genotyping-by-sequencing (GBS)



Optimization of GBS

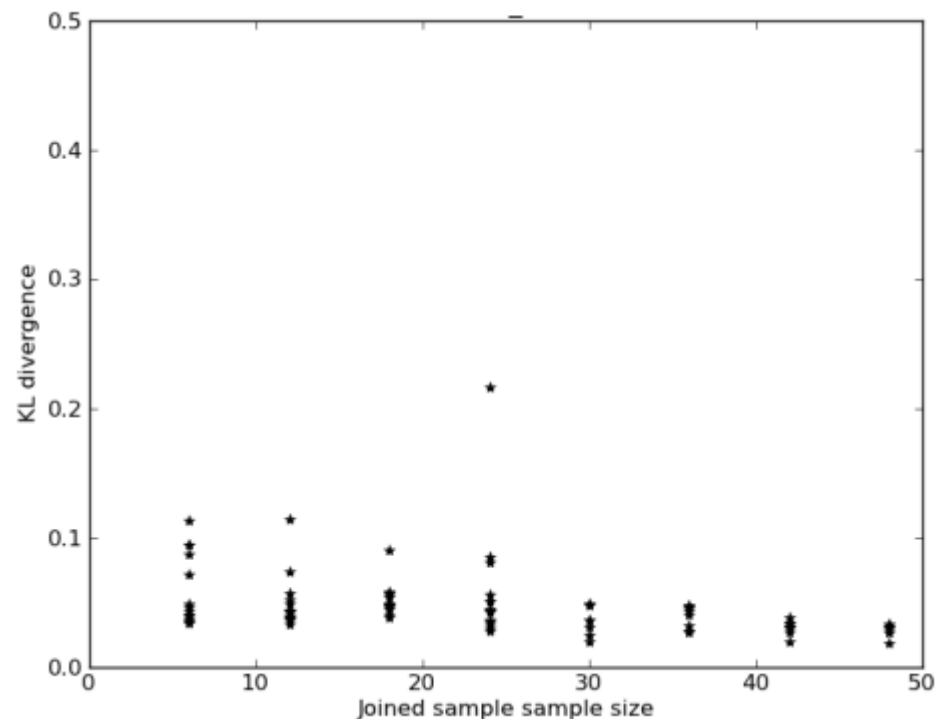
- We need to do this in a resource effective manner
- Can we sample multiple individuals effectively in one sample?
- How many individuals per accession do we need to sample to effectively represent the accession

GBS for landraces

- Highly repeatable GBS protocol optimized for maize samples developed at DArT

P0161

- Preliminary work indicated that this protocol was applicable to the use of bulks and methods developed to estimate allele f in mixed samples

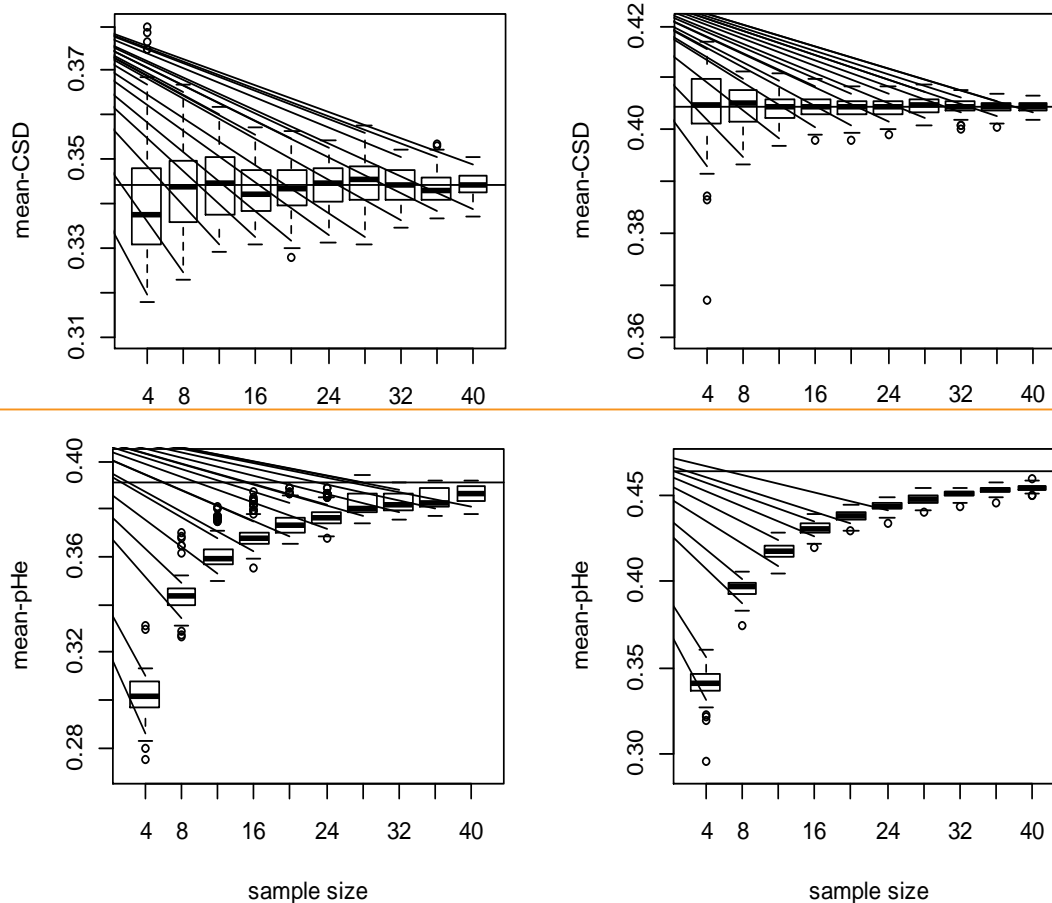


GBS for landraces- sampling

- 8 diverse accessions
- GBS of 92-96 individuals per accessions
- Simulated sampling of a defined number of individuals (4, 8, 12 etc.) from the dataset and determine distance and diversity indices of the population and samples

GBS for landraces- sampling

- What do you want to measure- distance or diversity ? both

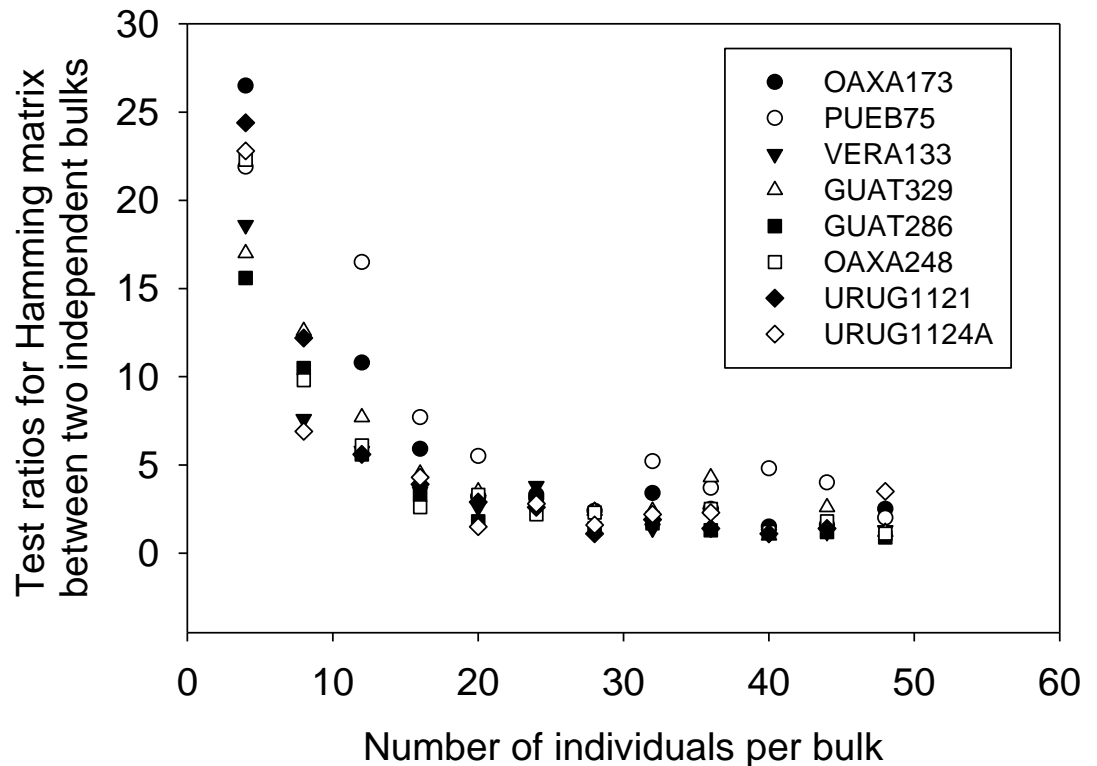


GBS for landraces- bulks

- 8 diverse accessions
- Extracted DNA from bulked tissue of individuals; 4, 8, 1248
- Created two independent bulks per sample size per population

GBS for landraces- sampling

- Divergence between bulks stabilizes from 20 for all accessions

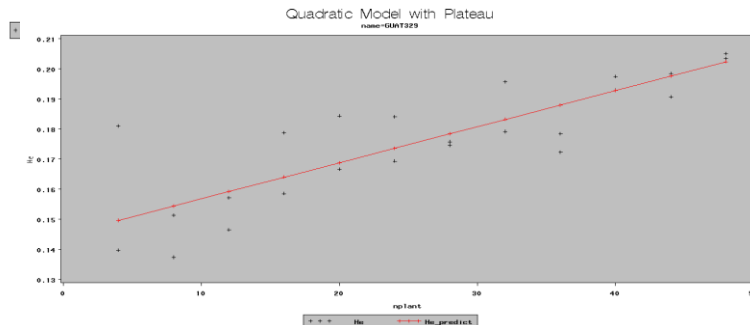
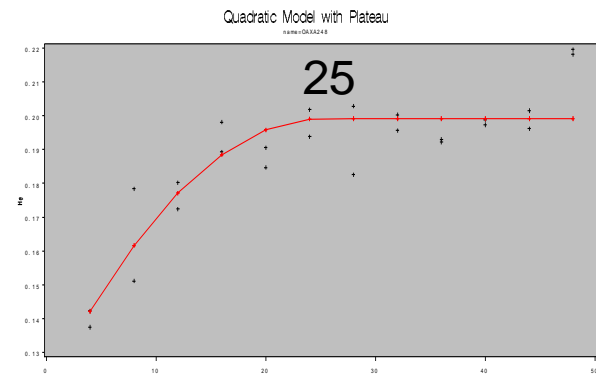
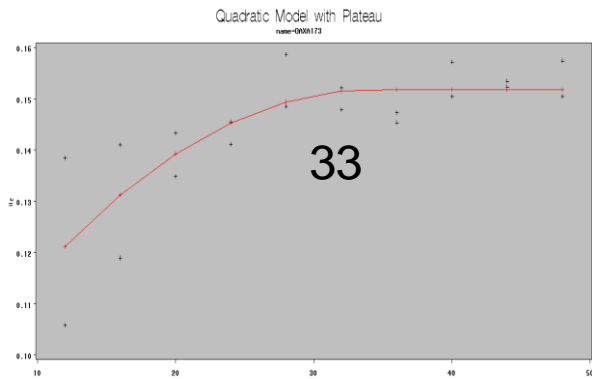
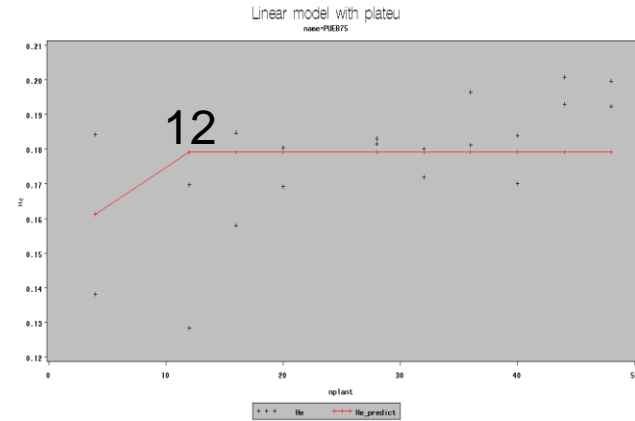
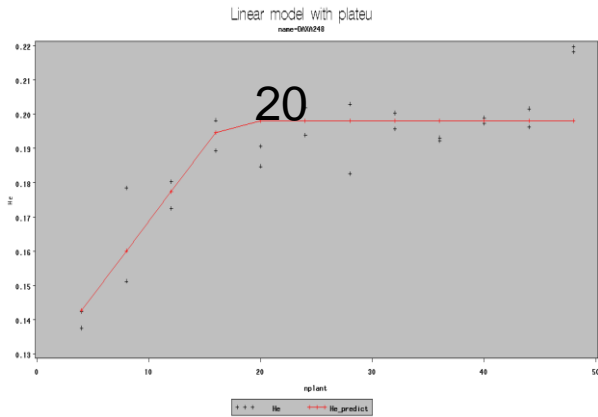


GBS for landraces- sampling

- Is there a threshold in diversity measures?
- Calculated numerous diversity indices
- Linear and quadratic regression using segmented model with plateau to investigate influence of bulk size on diversity measures

GBS for landraces- sampling

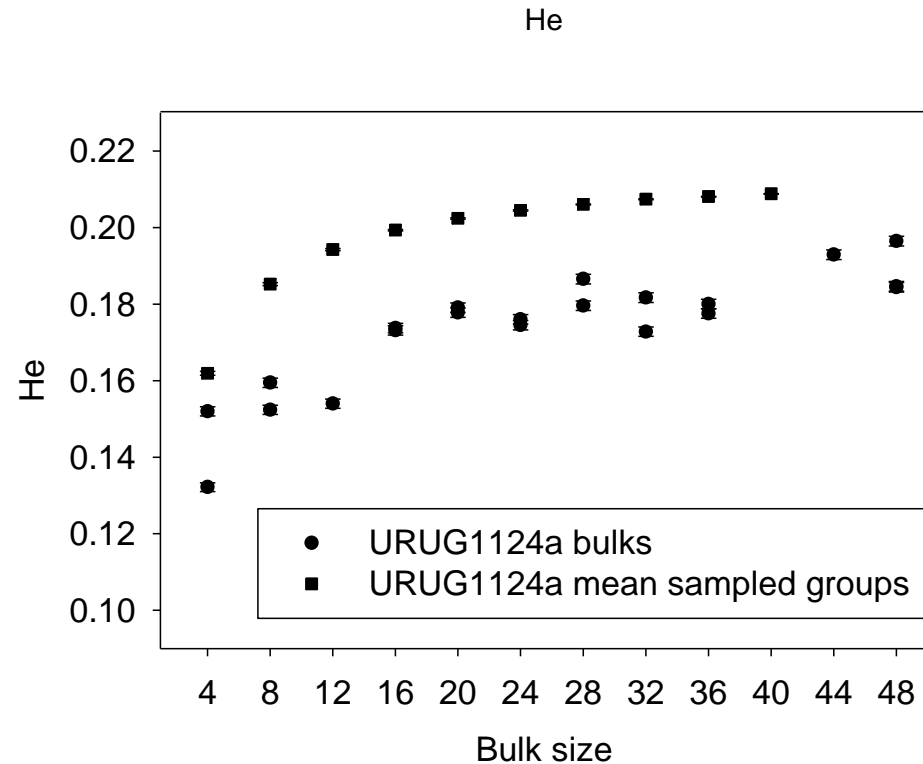
Mean He



12-40+

GBS of bulk samples

- Potentially some resolution is lost considering diversity measures using bulks
- Further analysis conducted to address this



GBS of bulk samples

- Recent analysis (yesterday)
- Comparing the effect of bulk v individual on estimation of H_e (bulks v component individuals)- bulks result in a slight overestimation of diversity

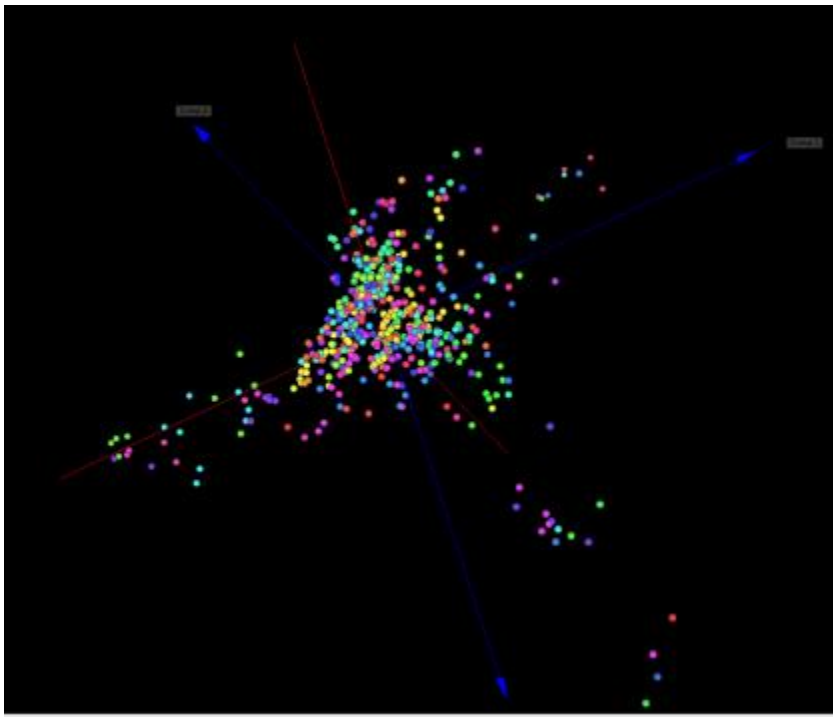
GBS of bulk samples

- Can we genotype bulks and derive population level allele frequencies?
- Yes and these are representative of population level genetic distances
- Thirty plants per accession provides sample with little divergence of GBS profile from individuals, facilitates use of one bulk per accession, optimization of diversity estimates, is practically achievable,
- Some overestimation considering diversity measures – estimates of effect determined

GBS of maize landraces

- Started to implement a bulked GBS approach- 1500 accessions
- Processing 12,000
- Landrace population level GBS profiles \$40/accessions
- Thirty plants per accession (equivalent cost \$1200 per accession)
- Some loss of resolution/over estimation - document errors in estimates

Towards a Maize molecular atlas



Thank you

