ABSTRACT: 352

DIVERSITY AND ASSOCIATION MAPPING IN AN IITA SOYBEAN PANEL

AGRAMA H[1*], BELZILE F[2], BOAHEN S[3], DION P[2], TEFERA H[1] AND SONAH H[2]

[1] *International Institute of Tropical Agriculture (IITA), IITA-Zambia, Lusaka, Zambia*
[2] *Département de phytologie, Pavillon Charles-Eugène Marchand, 1030, avenue de la Médecine, Université Laval, Québec, Canada*
[3] *International Institute of Tropical Agriculture (IITA), IITA-Mozambique, Nampula, Mozambique*
E-mail: H.Agrama@cgiar.org

Understanding the basis of complex quantitative traits of economic importance is a major tactic behind the progress in plant breeding. During the past decade, association mapping, commonly applied in human and animal genetics, has shown greater promise and power of mapping complex quantitative traits in plants compared to traditional linkage mapping. Germplasm diversity is the mainstay for crop improvement and genetic dissection of complex traits. A collection of 200 IITA soybean (*Glycine max*) lines were analyzed by genotyping by sequencing to identify SNP markers associated with nodulation traits. Understanding genetic diversity, population structure, and the level and distribution of linkage disequilibrium in target populations is of great importance and a prerequisite for association mapping. In this study, a genome-wide association analysis was conducted to detect key single-nucleotide polymorphisms (SNPs) associated with promiscuous nodulation in soybean. Genome-wide association mapping using models controlling both population structure (Q) and relative kinship (K) were performed to identify the marker loci/QTLs underlying the naturally occurring variations of nodulation in an IITA soybean panel. These significant SNPs will help to better understand the genetic basis of nodulation-related physiological traits, and facilitate the pyramiding of favorable alleles for nitrogen fixation traits in soybean marker assisted selection schemes for high nodulation efficiency. Efforts to identify novel variability in these genes may present opportunities to improve soybean natural nodulation quality.

Diversity and association mapping in an IITA soybean panel

Agrama H[1], Belzile F[2], Boahen S[3], Dion P[2], Tefera H[1] and Sonah H[2]

[1] International Institute of Tropical Agriculture (IITA), IITA-Zambia, Lusaka, Zambia

[2] Département de phytologie, Pavillon Charles-Eugène Marchand, 1030, avenue de la Médecine, Université Laval, Québec, Canada

[3] International Institute of Tropical Agriculture (IITA), IITA-Mozambique, Nampula, Mozambique

Why soybean were considered by IITA

- Little effort in improving soybean in Africa
- Yield was extremely low
- Low seed viability
- Poor nodulation with native *Rhizobium*
- High shattering in the moist and dry savanna

Importance of soybean in Africa

- Soybean constitutes an important component of the smallholder cropping systems in Africa
  - enhancing household food and nutrition security
  - holds considerable potential for arresting soil fertility decline
  - raising rural incomes and reducing poverty

Why breeding for promiscuous nodulation?

- Breeding for promiscuous nodulating genotypes was one of the approaches IITA followed to enhance biological nitrogen fixation of tropical soybeans
- Soybeans that nodulate effectively with diverse indigenous rhizobia are considered as promiscuous, and the characteristic promiscuity (Kuneman et al., 1984).
- Hence, promiscuous genotypes of soybean form symbiotic association with available *Rhizobium* strains in the soil and hence fix atmospheric nitrogen whilst non-promiscuous genotypes need specific rhizobial strains to fix nitrogen from the air

Generally

- Soybean varieties developed for promiscuous nodulation with the indigenous rhizobia
  - were considered to increase production of soybean in tropical Africa
  - with minimum cost affordable to small-scale farmers

Materials and Methods

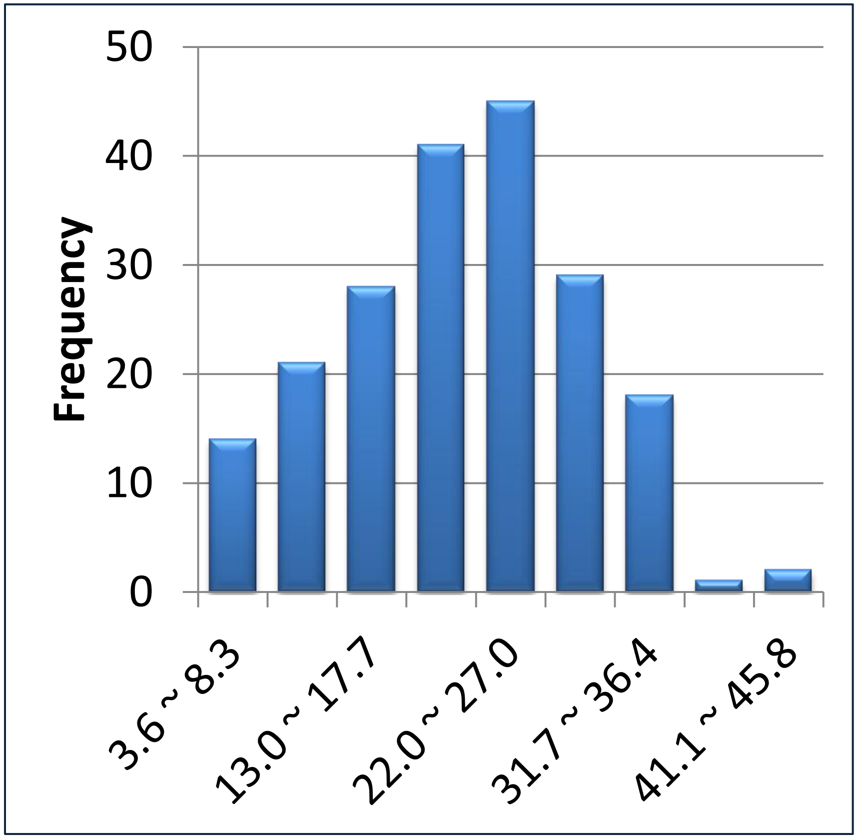
- Germplasm diversity is the mainstay for crop improvement and genetic dissection of complex traits
- A collection of 280 IITA soybean (*Glycine max* L) lines and 15 released varieties were evaluated for nodulations
- Two locations Malawi and Mozambique in two seasons 2010 and 2011
- The accessions were analyzed by genotyping by sequencing to identify SNP markers associated with nodulation traits

Nodulating Characterization

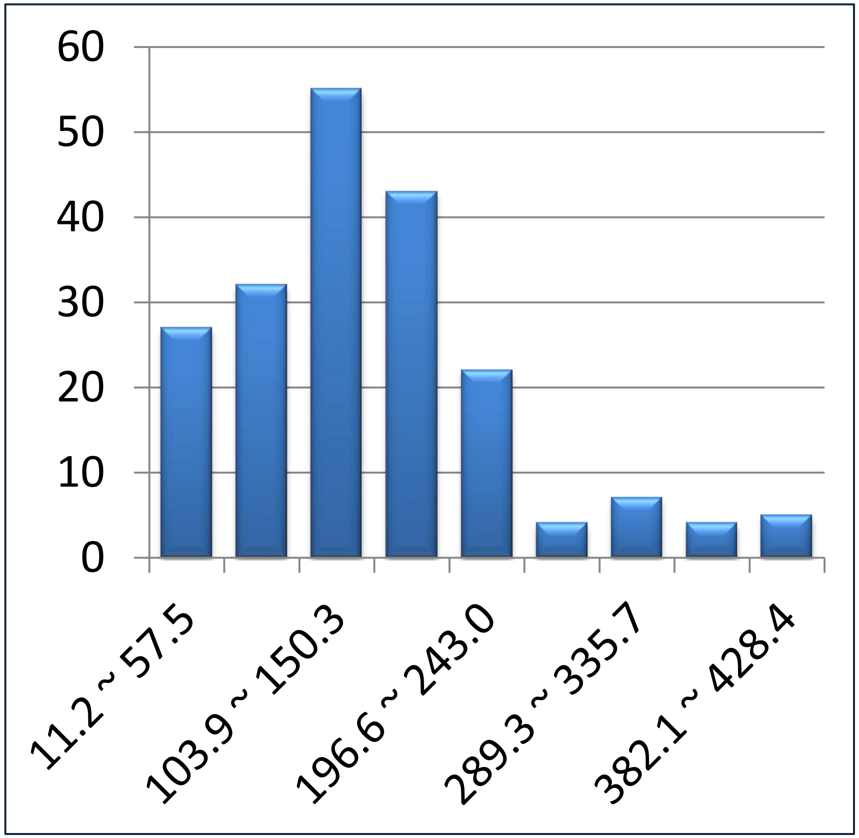
- The roots of each of 20 plants 45 days old were washed and carefully rinsed with water
- The nodules were harvested
  - counted (NodNo), then oven dried at 65^{o}C for 4 days
  - weighed to determine the total Nodule dry weight *mg* (NoddryWt) per plant.

Phenotypic distribution of two nodulation traits

No of nodules/plant



Nodule dry weight *mg/plant*



Association Mapping

Population-based genetic association study

- A promising alternative approach turns
  - from families to populations,
  - and from linkage analysis to association studies

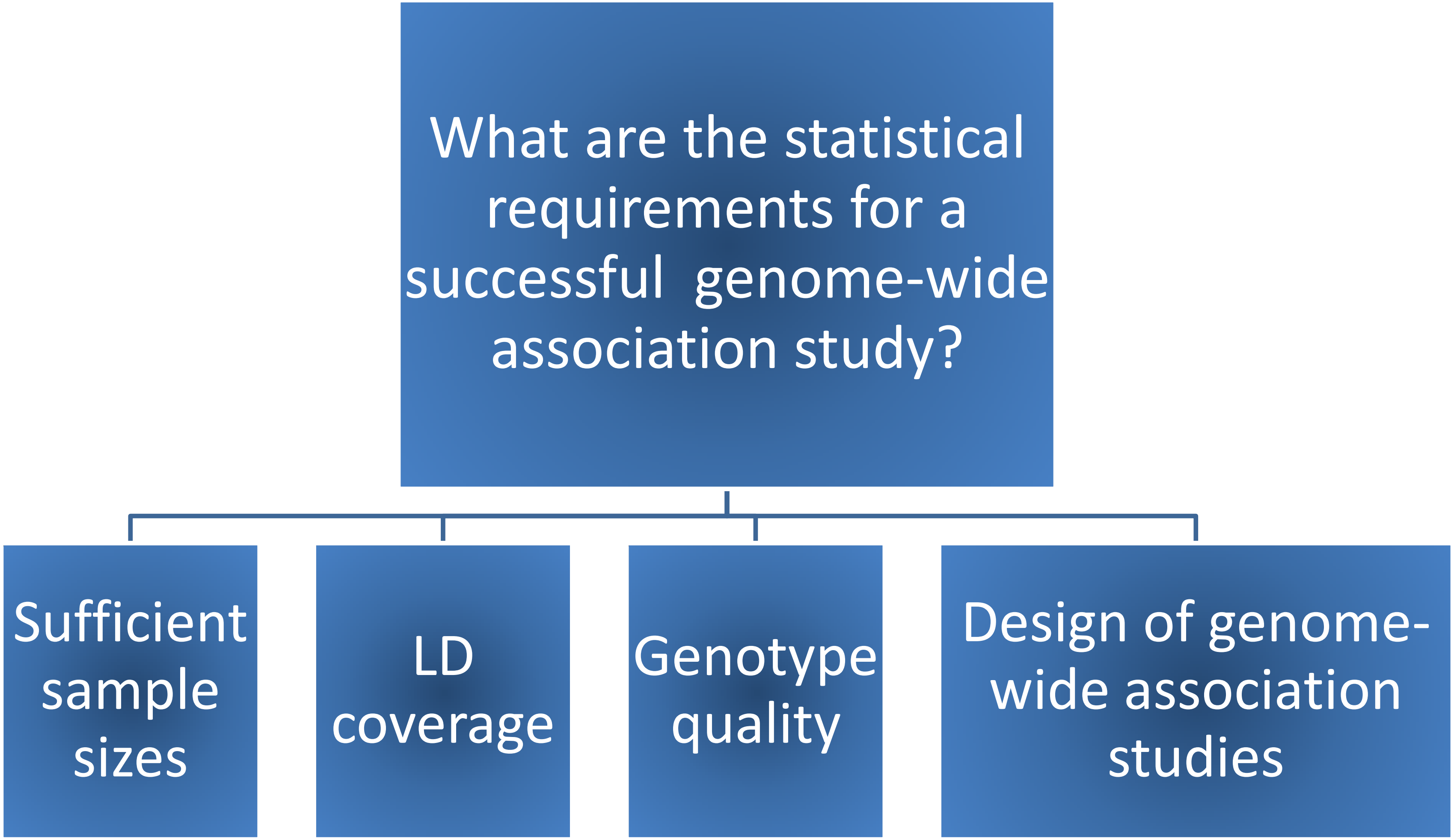
Association Mapping

- Association Mapping can be conducted directly on the breeding material, therefore:
  - Direct inference from data analysis to breeding is possible
  - Relevant genetic background effects are sampled
  - Phenotypic variation is observed for most traits of interest
  - Marker polymorphism is higher than in biparental populations
  - Routine variety trial evaluations provide phenotypic data over years and locations

The principle of association mapping

- Utilize ancestral recombination events in natural populations to make marker–phenotype associations
- Evaluate whether certain alleles within a population are found with specific phenotypes more frequently than expected

Association Mapping: Overview



What are the statistical requirements for a successful genome-wide association study?

- Sufficient sample sizes
- LD coverage
- Genotype quality
- Design of genome-wide association studies

Association Mapping: Overview

For association mapping, a given sample may contain either

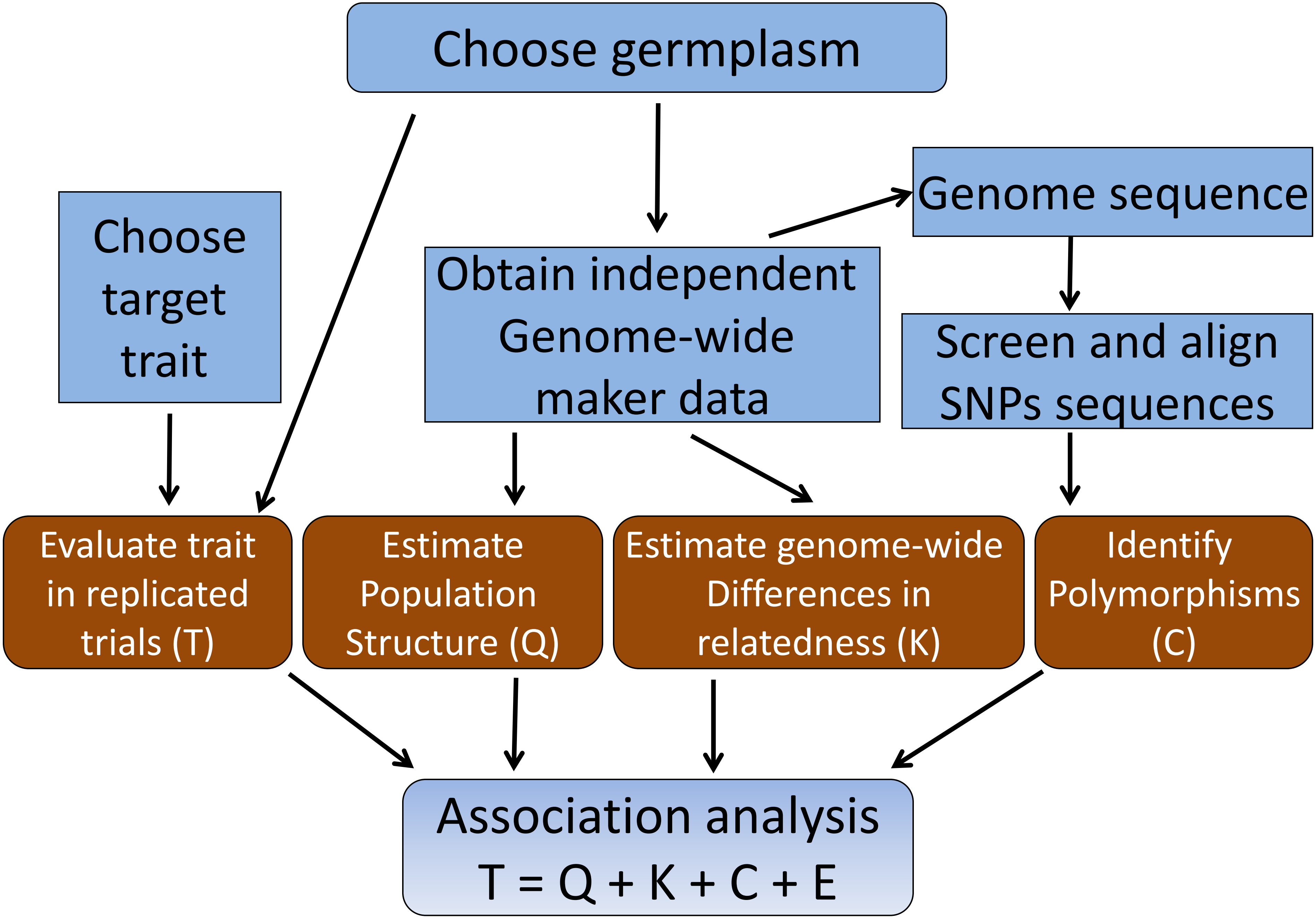
population structure

(associated with local adaptation or diversifying selection),

or familial relatedness

(from recent co-ancestry),

or both



Choose germplasm

Choose target trait

Obtain independent Genome-wide maker data

Genome sequence

Screen and align SNPs sequences

Evaluate trait in replicated trials (T)

Estimate Population Structure (Q)

Estimate genome-wide Differences in relatedness (K)

Identify Polymorphisms (C)

Association analysis

$$T = Q + K + C + E$$

Genotyping

- One promising approach is genotyping-by-sequencing (GBS) which uses enzyme-based complexity reduction (using restriction endonucleases to target only a small portion of the genome) coupled with DNA barcoded adapters to produce multiplex libraries of samples ready for NGS

Elshire et al., 2011; Poland & Rife 2102

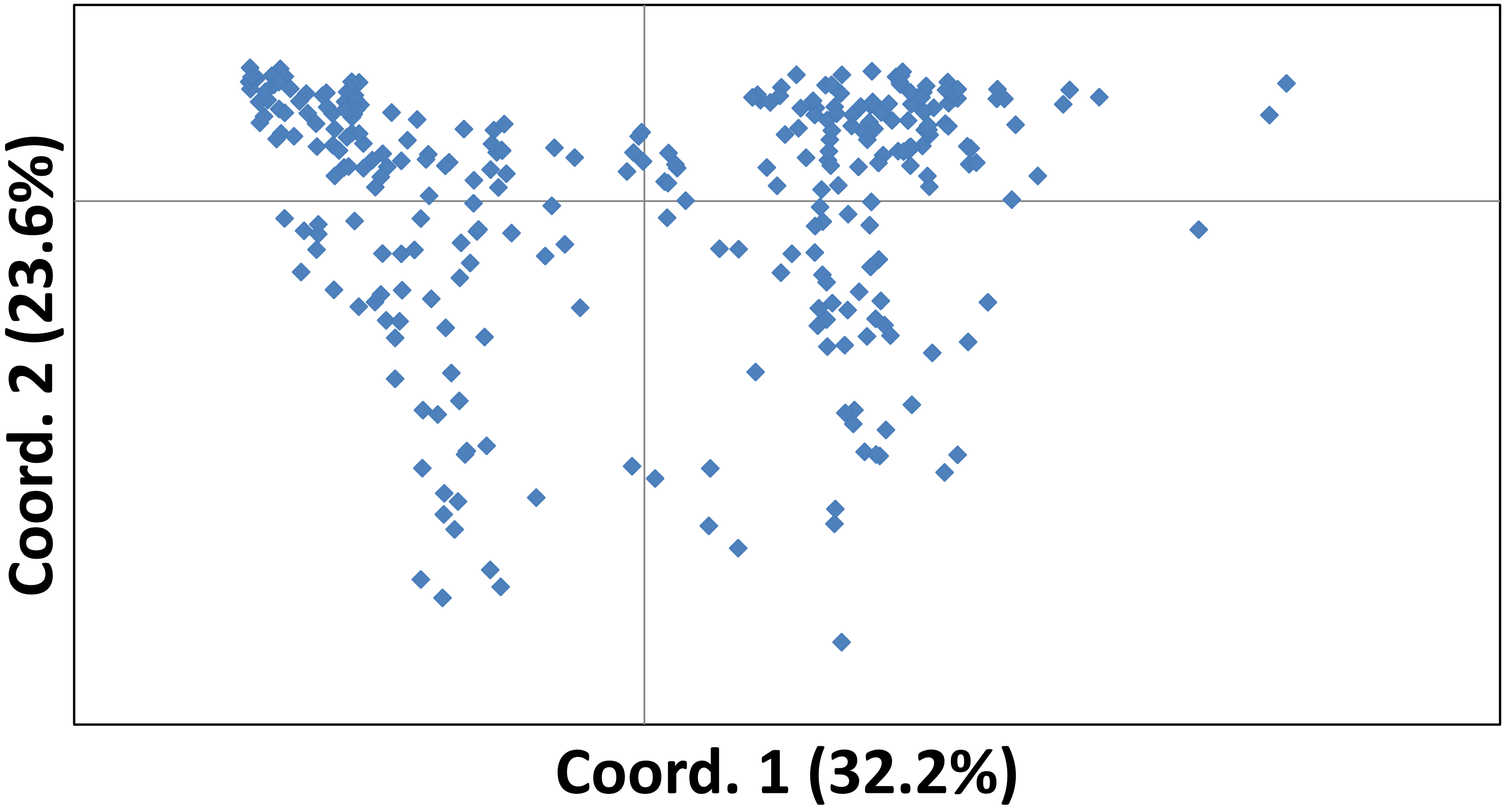
Genotyping

- Here we show that GBS can be applied directly to breeding programs and produce *de novo* molecular markers suitable for whole-genome association.
- We applied GBS to the set of advanced soybean breeding lines from the International Institute of Tropical Agriculture (IITA).

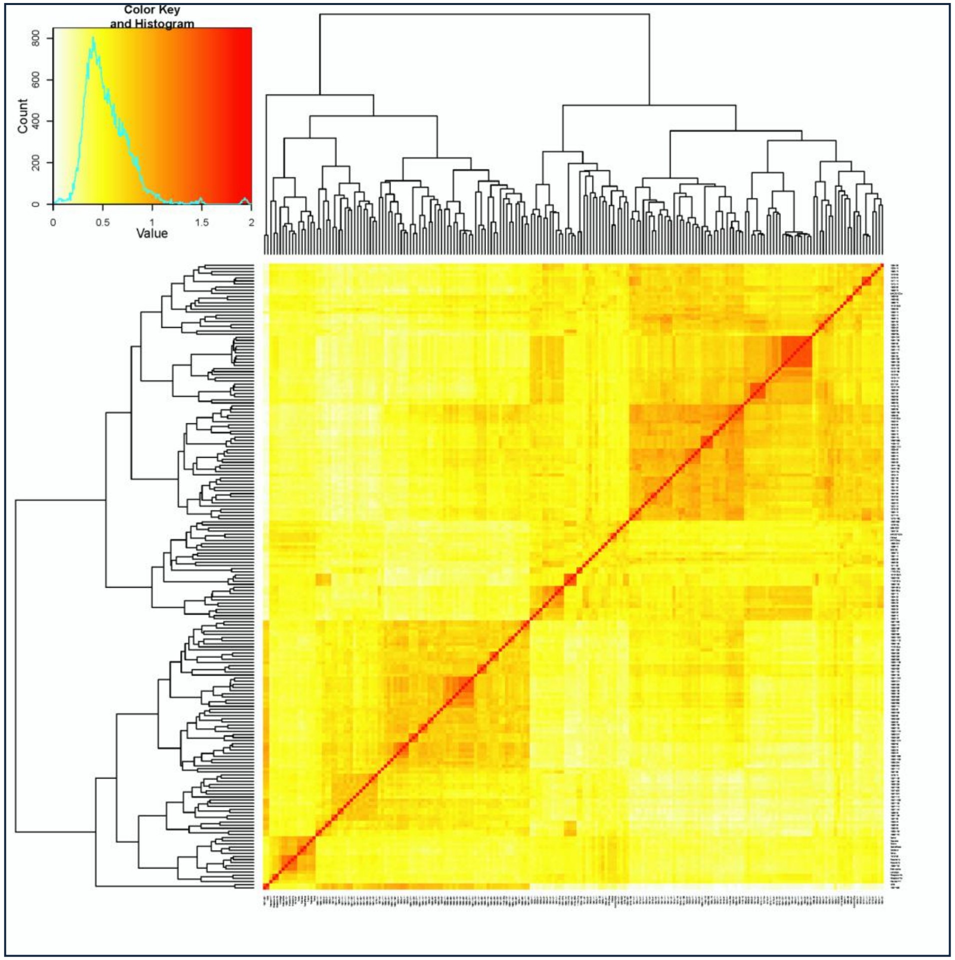
Analysis

- Number of SNPs: 4974
- Population structure among soybean accessions was performed using the Principal Coordinate Analysis (PCA)
- Association between SNP alleles and different nodulation traits data was performed with GAPIT (Genomic Association and Prediction Integrated Tool), Lipka et al. 2012

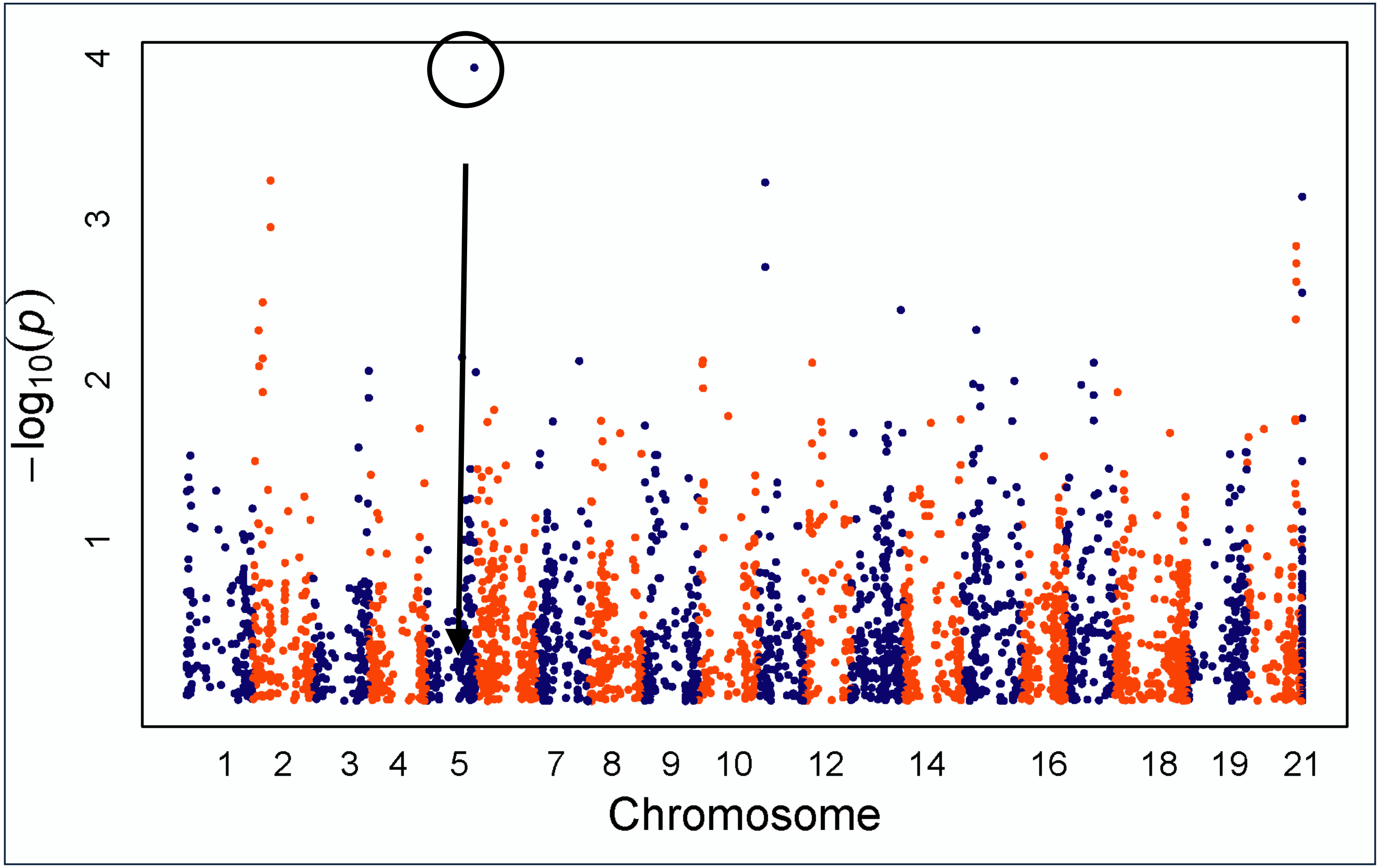
Principal Coordinates



A heat map of the values in the values in the kinship matrix

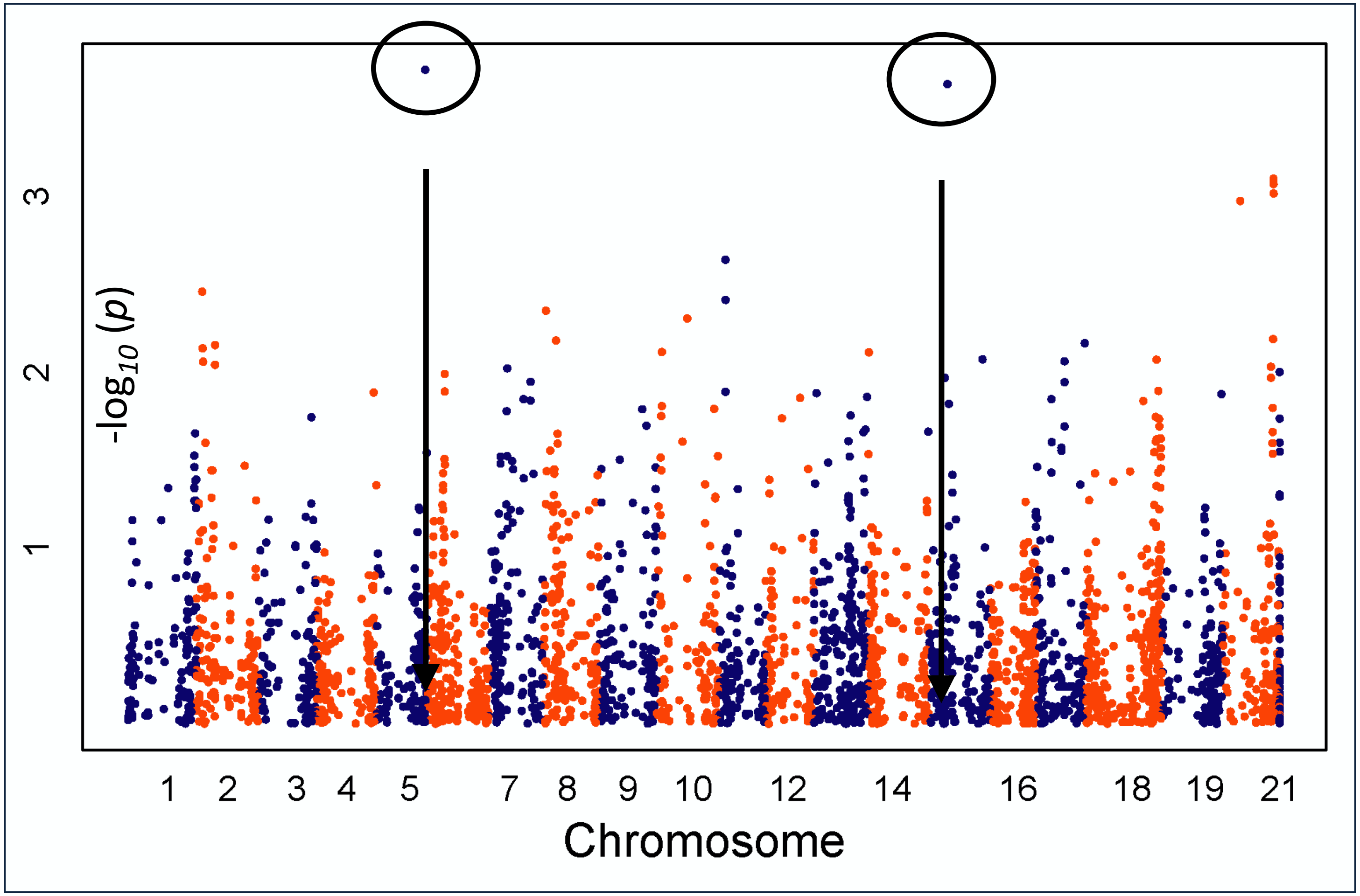


Manhattan plot: SNPs with stronger associations with the No of nodules/plant



The X-axis is the genomic position of the SNPs in the genome, and the Y-axis is the negative log base 10 of the P-values. The chromosomes are alternatively colored.

Manhattan plot: SNPs with stronger associations with the Nodule dry weight *mg*/plant



The X-axis is the genomic position of the SNPs in the genome, and the Y-axis is the negative log base 10 of the P-values. The chromosomes are alternatively colored.

Conclusions

- Rapid advances in output from NGS platforms with corresponding decreases in cost have made sequence-based genotyping a very attractive and practical approach to rapidly characterize genomes and populations

Future considerations

Soybean is highly nutritious and adaptable

it offers resources to address world food issues through current and future utilization practices

Soybean production is expected to increase

increase in proportion to increased demand

Soybean has potential to improve

with application of newer genomic technologies

Challenges and threats to production

Abiotic limitations

- weather related phenomena
- soil nutrient availability
- salinity
- response to photoperiod

Biotic limitations

- pathogens
- pests
- weeds

Challenges and threats to production

The implementation of molecular markers closely associated with desirable traits is being used to increase the efficiency and effectiveness of conventional breeding by indirect selection of the desirable plants in segregating population

- improving grain yield
- biological nitrogen fixation
- pod shattering
- diseases — rust
- resistance to lodging
- tolerance to low Phosphorus
- drought tolerance
- dual-purpose soybeans

Thanks