Genotyping-by-sequencing reveals population structure in a Napier grass (*Pennisetum purpureum*) collection and identifies marker-trait associations underpinning the selection of improved varieties

I. Introduction

Napier grass (*Pennisetum purpureum* L.) is an important tropical forage used mainly as a cut and carry feed. Important attributes include high yield/unit area, year-round availability due to its perennial nature, resistance to most diseases and pests, ease of establishment and propagation, and fast regrowth. Napier grass genetic characterization to date has mainly relied on assessing phenotypic traits and using molecular markers such as Amplified Fragment Length Polymorphisms (AFLPs) and microsatellites (SSRs). Recent advances in genotyping by sequencing (GBS) approaches provide a cost effective method of identifying genome-wide molecular markers in species with limited genomic information. These markers can be used to detect regions of the genome controlling important agronomic traits within germplasm collections by applying genome-wide association studies (GWAS). Here we report on a diversity analysis of the ILRI genebank Napier grass collection using the GBS method of the DArTseq platform, that combines genome complexity reduction using restriction enzymes and next-generation sequencing (NGS). The analysis revealed a substantial amount of genetic variation. We also present a preliminary association analysis of markers with agronomic and morphological traits using historical data.

II. Napier grass population

The ILRI forage genebank holds a diverse set of Napier grass accessions, collected across a range of environments and origins, and maintained in Ethiopia at their Bishoftu and Ziway sites. The ILRI genebank has also acquired distinct lines from EMBRAPA (Brazilian Agricultural Research Corporation). Ninety-four accessions were selected from the collections for further study.

III. Genotyping-By-Sequencing (GBS)

The nighty-four accessions were subjected to genotyping by GBS, using the DArTseq platform. The genotyping produced 148,493 genome wide SNP and Silico DArTseq markers with an average call rate of 83% and 96%, respectively. Figure 1 shows the distribution of polymorphic information content (PIC) of the markers.
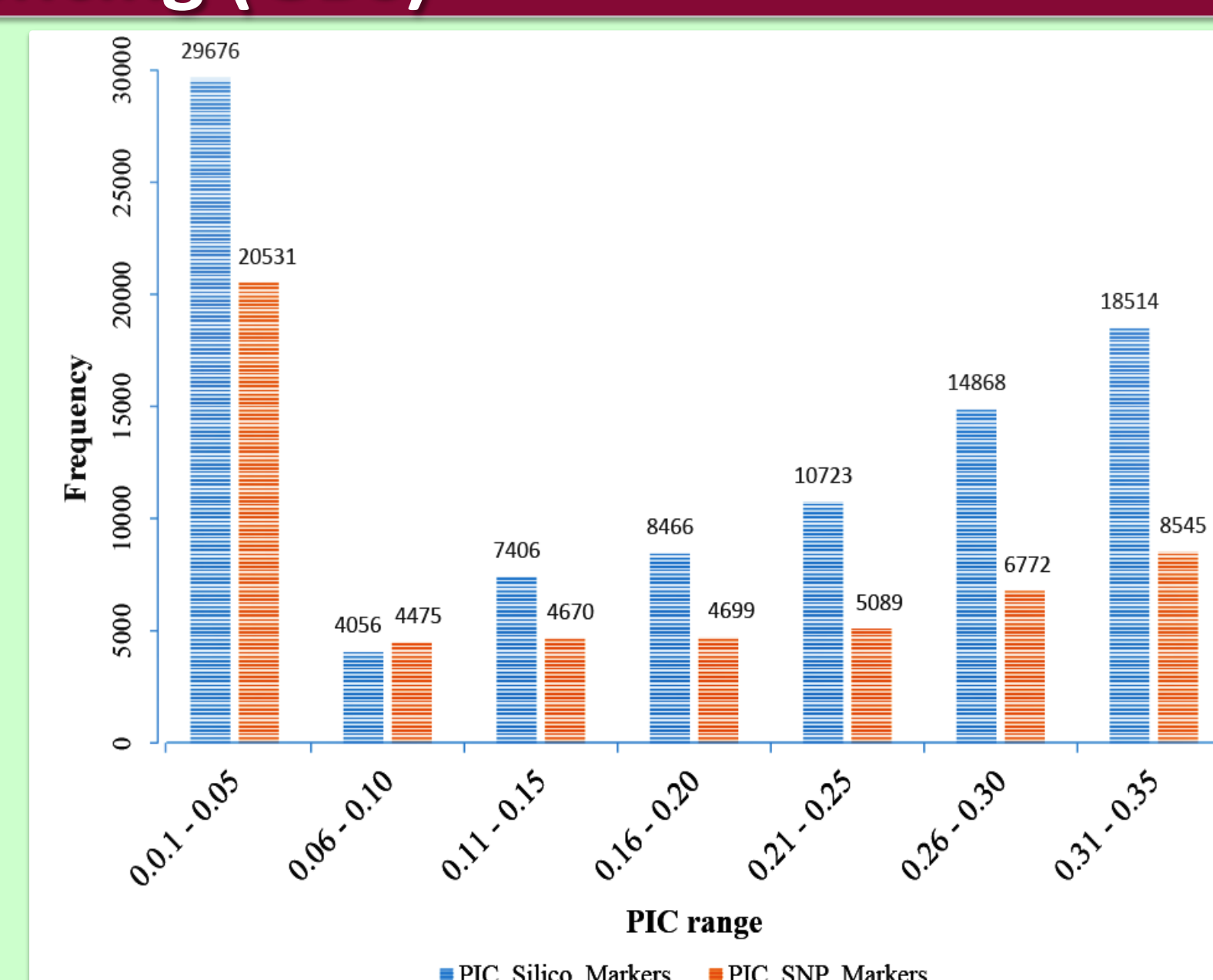

Figure 1. Distribution of polymorphic information content (PIC) of Silico (blue) and SNP (orange) markers

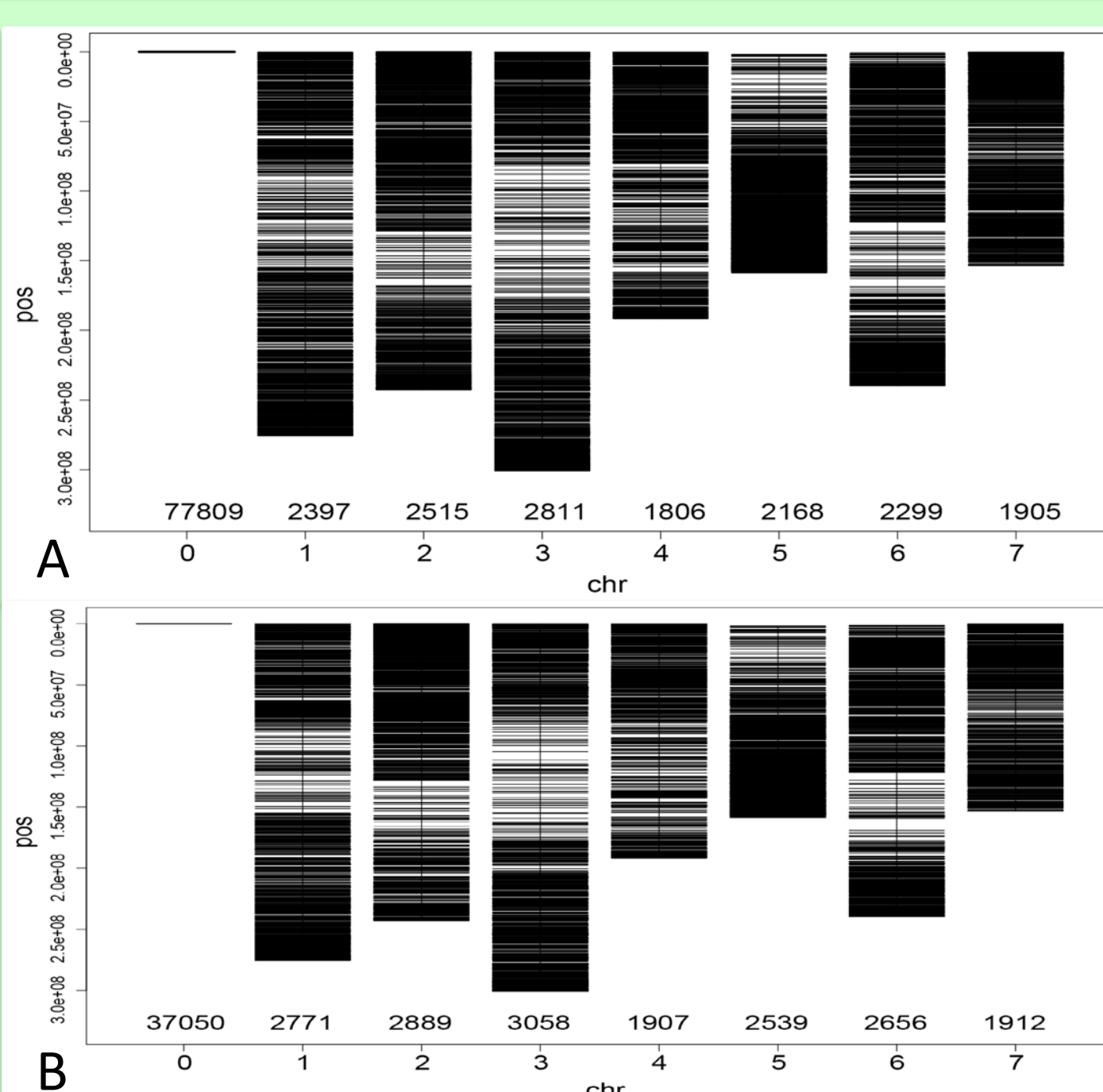## IV. Physical map position of markers

The genome wide markers were mapped on the genome of a related species, pearl millet (*P. glaucum*). The map information was important to select a few representative markers with high-resolution power for diversity analysis and also to run marker-trait association analysis and to predict genomic regions associated with important agronomic and morphological traits.

Figure 2. Genome wide distribution of Silico (A) and SNP (B) markers. The markers that were not mapped indicated by 0.

V. Diversity in Napier grass population

Diversity and population structure in the Napier grass collection was analyzed using a selection of 1,000 highly polymorphic markers. The presence of up to 4 sub populations, in which three main subpopulations are more clearly demarcated, was observed by STRUCTURE, PCA and phylogenetic analysis. Group I consists of mainly EMBRAPA material and group III is mainly ILRI material with groups II and IV containing material from both collections.
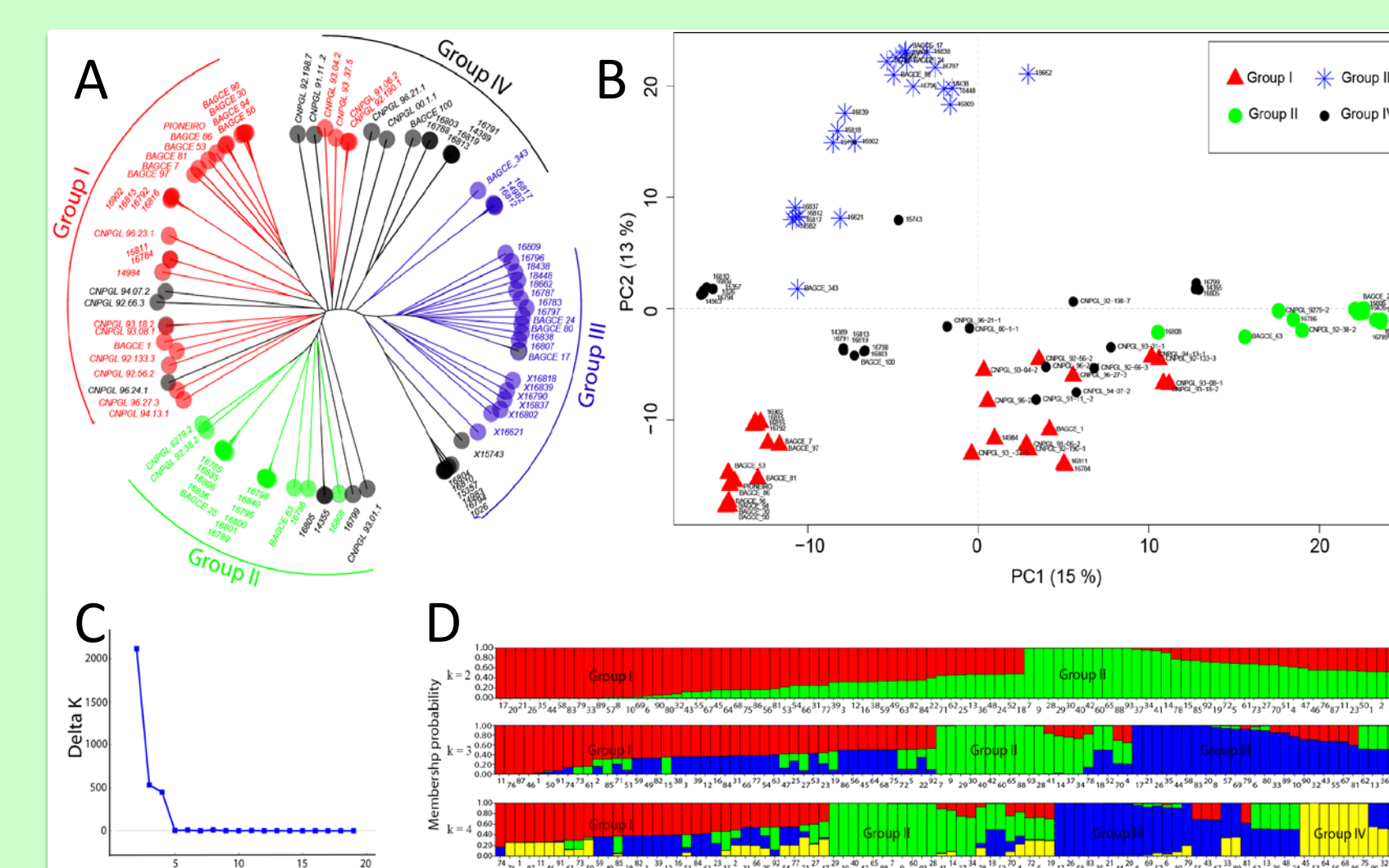

Figure 3. Clusters of the 94 Napier grass accessions. (A) NJ tree (B) PCA (C) The delta K suggesting two major population and up to 4 sub populations (D) Bar plots based on the admixture model in STRUCTURE. The colors are according to the STRUCTURE k = 3.

VI. Historical phenotype data in Napier grass population

Figure 4. Correlation of traits

Historical phenotypic data collected on 47 accessions from the collection in 1993/94 for the below traits were used for the selection of best bets and for marker-trait association analysis

P.Height = plant height
Sthik = stem thickness
Inno = internode length
TSDW = total stem dry weight
TLDW = total leaf dry weight
Yield

VII. Markers associated with the traits

Marker-trait association analysis using the historical agronomic and morphological data on 47 accessions detected a total of 13 molecular markers associated with the 'plant height' trait, which is highly correlated with 'stem dry weight' and 'yield'.
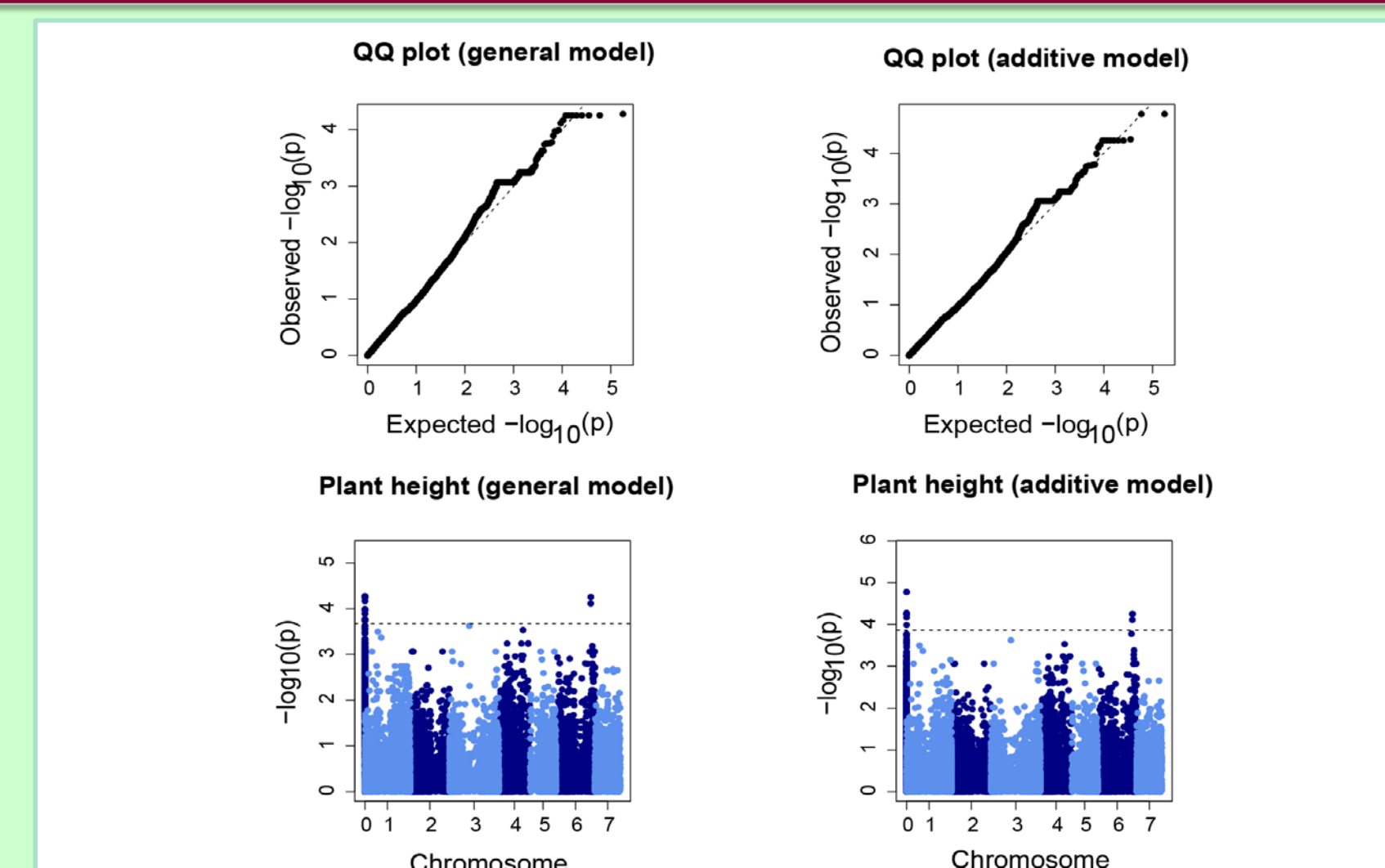

Figure 5. Manhattan plots showing the association of markers with the plant height trait, using the general and additive models. Quantile-quantile (QQ) plots showed for each model

VIII. Conclusions

Genotyping using the DArTseq platform generated high-density markers with a reasonable distribution across the genome which were suitable for diversity and marker-trait analysis. The diversity analysis revealed the existence of a substantial amount of variation in the ILRI genebank Napier grass collection and identified some unique materials from the EMBRAPA collection, showing the suitability of the population for further genetic study. The associated markers will be used for the selection of higher yielding genotypes and applied in breeding programmes for the development of improved varieties after being verified on a larger population in a replicated field trial.

Meki S. Muktar, Jean Hanson, Abel Teshome, Alemayehu Teressa Negawo, and Chris Jones

Feed and Forage Development program, International Livestock Research Institute (ILRI), Box 5689, Addis Ababa, Ethiopia

ILRI INTERNATIONAL LIVESTOCK RESEARCH INSTITUTE

CGIAR