High-Density Molecular Characterization and Association Mapping in Ethiopian Durum Wheat Landraces Reveals High Diversity and Potential for Wheat Breeding











Mengistu D.K.<sup>1,2</sup>, Kidane Y.G.<sup>1,3,4</sup>, Catellani M.<sup>1</sup>, Frascaroli E.<sup>5</sup>, <u>Fadda C.<sup>4</sup></u>, Pè M.<sup>1</sup>, Dell'Acqua M.<sup>1</sup> <sup>1</sup> Scuola Superiore Sant'Anna di Pisa, Italy; <sup>2</sup> Mekelle University, Ethiopia; <sup>3</sup> Sirinka Agricultural Research Center (ARARI), Ethiopia; <sup>4</sup> Bioversity International, Ethiopia; <sup>5</sup> University of Bologna, Italy

## Abstract

Durum wheat (*Triticum turgidum* subsp. *durum*) is a key crop worldwide, yet its improvement and adaptation to emerging environmental threats is made difficult by the limited amount of allelic variation included in its elite pool. New allelic diversity may provide novel loci to international crop breeding through quantitative trait loci (QTL) mapping in unexplored material. We test **81,587** markers scoring **30,155** single nucleotide polymorphisms and use them to survey the diversity, structure, and genome-specific variation in a representative panel of **300 Ethiopian durum wheat landraces** and a siding collection of **Mediterranean durum wheat accessions**. We phenotype the Ethiopian panel for **ten agronomic traits** in two highly diversified Ethiopian environments for two consecutive years, and use this information to conduct a **genome wide association study**. We identify several loci underpinning agronomic traits of interest, both confirming loci already reported and describing new promising genomic regions. These loci may be efficiently targeted with molecular markers already available to conduct marker assisted selection in Ethiopian and international wheat. We show that Ethiopian durum wheat represents an important and **unexplored source of durum wheat diversity**.



**Figure 1** – Altitudinal map of Ethiopia, green to red increasing elevation. Sampling points, when available, are marked as blue dots. Landraces have been sampled in 16 regions (only macro regions reported in map). Light blue points mark experimental fields (see Genome-wide association)



#### **Diversity panel assembly**

The plant material was collected at the Ethiopian Biodiversity Institute (EBI; www.ibc.gov.et). Passport data were a necessary requirement to consider accessions. The panel is composed of accessions recorded as durum and/or tetraploid wheat collected in 16 different regions of Ethiopia. We used a spike derived from a single plant per landrace as a seed donor for growing the seedlings that provided the plant material for subsequent DNA extraction and eventually for field trials. The Ethiopian durum wheat collection is composed of 287 landraces representative of durum wheat variation in the region (Fig. 1) and 24 improved varieties approved for cultivation in Ethiopia, virtually all improved material cultivated. The Ethiopian dataset is sided by 38 Mediterranean durum wheat accessions and lines representing the international allele pool.

### Genotyping

Of the **81,587** probes available on the chip, 2,899 consistently failed. Average heterozygosity in the Ethiopian durum wheat dataset was equal to 0.02, and failure rate equaled 0.15. As failure rates increased, so did heterozygosity. Both heterozygosity and failure rates were consistently higher in Ethiopian than in Mediterranean durum wheat, especially in the top outliers (Fig. 2). Probes failing in more than 25% of the samples were removed as possible indicators of markers with a low specificity. The array scored **30,155** polymorphic loci in the Ethiopian durum wheat subset. The same array scored **21,069** polymorphic SNPs when considering Mediterranean material alone.



**Figure 2** – Heterozygosity (a) and failure rate (b) are higher in Ethiopian material (land, landraces; imp, improved) than in Mediterranean varieties (Med), suggesting the presence of substantial unobserved variability.



# Genetic diversity

We performed a PCA on the SNPs data of Ethiopian and Mediterranean durum wheat samples (Fig. 3). PC1 and PC2 accounted for 20.1% and 4.3% of the variance, respectively. Three main findings emerge: i) Ethiopian landraces are clearly separated from the international allele pool, ii) Ethiopian landraces collect a broader diversity than international material, and iii) the improved varieties approved for cultivation in Ethiopia cluster together with Mediterranean accessions. Some samples do not follow this general rule, possibly because of contamination at the seedbank

level. When improved varieties were removed from the analyses, landraces revealed

in a **Structure analysis** the existence of **10 genetic clusters** (termed Q1 to Q10)

(Fig. 4). Despite clearly distinct sampling points, individuals are widely admixed, and

very few "pure lines" can be found. A substantial exchange of germplasm has

occurred. Amhara South Gonder (AM\_SG) and Amhara West Gonder (AM\_WG),

although close to each other, are mostly composed of Q4 and Q2 clusters,

respectively. This may be contributed either by local adaptation or by barriers to seed

**Figure 3** – PCA space for molecular diversity. Ethiopian landraces (green) are completely secluded from the Mediterranean allele pool (red), and much more diverse on PC2. Improved lines in Ethiopia (blue) make no use of local alleles. Low PC loadings suggest little structure.

#### **Genome-wide association (GWA) study**

Phenotypes were measured for 2 years, 2 replicas in 2 locations (Fig.1). Days booting (DB), flowering (DF) and maturity (DM), number of effective/productive tillers per plant (NET), plant height (PH), spike length (SPL), number of seeds per spike (SPS), grain yield (GY), above ground biomass (Bm), and thousand grain weight (TGW). We found significant (p<0.001) variability due to the genotypes for all traits. For combined data the broad sense heritability estimated for the traits ranged from 0.25 for NET to 0.89 for DB (GY and GY components heritability greater than 0.5). Linkage disequilibrium (LD) decays fast in the collection (Fig. 5). Homoeologous chromosomes show different LD decay patterns. The GWA scan on phenology reported four major loci with alternating effects (Fig. 6 and Fig. 7). Sanna et al. (2014) reported flowering signals in similar positions on the homoeologous Chr 4B and Chr 7B. Structural variation may play a relevant role in Ethiopian durum wheat diversity. Long range LD may be seen at different genetic positions (Fig. 7), suggesting mismatch in positions as compared to the international genetic map (Maccaferri et al., 2015).

exchange (*e.g.* language, connecting roads, flour usages)



**Figure 5** – Linkage disequilibrium (LD) decay as a function of genetic distance in genome A (a) and genome B (b). Measured data is fitted to Hille and Weir (1988). LD decays fast in both genomes, differently for homoeologous pairs some notably having slower decay.



**Figure 4** – Genetic structure (10 genetic clusters colored as in legend) is not spatially structured. Each sampling region is represented by a bar filled by each genetic cluster proportionally to its prevalence. Q2 and Q4 are the sole exceptions. SSNP is poorly sampled and thus not reliable.



### Way ahead

The characterization of this panel is the first milestone of a multipronged approach supported by the International Doctoral **Programme in Agrobiodiversity** at Scuola Superiore Sant'Anna, Pisa, Italy. Farmer communities are involved in every aspect of the project, providing data and knowledge whilst benefiting of germplasm distribution and development. 50 landraces harnessing traits of interest and maximizing diversity were used to produce a nested association mapping panel (McMullen et al., 2009) using the improved line Asassa as recurrent founder. The **durum wheat NAM** is currently at F6 and comprises 6,723 RIL divided in 47 families. This tool is at once the first quantitative exploration of Ethiopian durum wheat in breeding and a cutting edge QTL mapping tool available for collaborators. A concurrent crowd sourcing approach is redistributing the most preferred landraces to Ethiopian farmers communities, currently involving more than 2,000 farmers screened for landraces adoption and preference whilst measuring agronomic value in diffused fields as a function of climate during growing season



**Figure 6** – Manhattan plot for the GWA scan for phenologic traits using the SUPER method (Wang et al. 2014). in R/GAPIT (Lipka et al., 2012). Significance (y axis) is the negative Log of FDR. Significance thresholds are given for FDR 5% (green) and FDR 10% (gray). Peaks are clearly visible on Chr 1B, 2B, 4A with alternating effect. An association on Chr 7A emerges at maturity

**Figure 7** – Pairwise LD on the entire genome, divided by linkage group. Marker trait associations from GWA above FDR 5% are circled in colors according to legend. Note the patchy pattern of high LD on some chromosomes (notably 1A, 3B, 4B), It is possible that SNP markers developed on international material have a different map position in Ethiopian landraces.

Days to booting
Days to maturity
Plant height

Biomass
Seeds per spike

O Grain yield

R<sup>2</sup> Color Key

Number of effective tillers

**References:** Hill and Weir (1988) *Theor. Popul. Biol.*, 33, 54–78; Lipka et al (2012) *Bioinformatics*, 28, 2397–2399; Maccaferri et al (2015) *Plant Biotechnol. J.*, 13, 648–663; Mcmullen et al. (2009) *Science*, 345, 125178; Sanna et al (2014) *J. Exp. Bot.*, 65, 3177-3188; Wang et al (2014) *PLoS ONE*, 9, e107684.

Poster presented at Tropentag 2016, Wien, Austria, 18th-21st September 2016