Data-driven, participatory characterization of farmer varieties discloses teff breeding potential under current and future climates

Aemiro B. Woldeyohannes[& 1,2], Sessen D. Iohannes[& 1], Mara Miculan[1], Leonardo Caproni[1], Jemal S. Ahmed[1], Kauê de Sousa[4,5], Ermias A. Desta[2], Carlo Fadda[3], Mario E. Pè[1], Matteo Dell'Acqua[1]*

[1] Center of Plant Sciences, Scuola Superiore Sant'Anna, Pisa, Italy

[2] Amhara Regional Agricultural Research Institute, Bahir Dar, Ethiopia

[3] Biodiversity for Food and Agriculture, Bioversity International, Nairobi, Kenya

[4] Digital Inclusion, Bioversity International, Montpellier, France

[5] Department of Agricultural Sciences, Inland Norway University of Applied Sciences, Hamar, Norway

[&] Equal contributing authors

*Correspondence to: Matteo Dell'Acqua, Scuola Superiore Sant'Anna, Piazza Martiri della Libertà 33, 56127, Pisa, Italy. Email: m.dellacqua@santannapisa.it

Keywords: *Eragrostis tef*; neglected and underutilized crop species; adaptation genomics; participatory variety selection; smallholder farming

Abstract

In smallholder farming systems, traditional farmer varieties of neglected and underutilized species (NUS) support the livelihoods of millions of growers and consumers. NUS combine cultural and agronomic value with local adaptation and transdisciplinary methods are needed to fully evaluate their breeding potential. Here, we assembled and characterized the genetic diversity of a representative collection of 366 Ethiopian teff (*Eragrostis tef*) farmer varieties and breeding materials, describing their phylogenetic relation and local adaptation on the Ethiopian landscape. We phenotyped the collection for its agronomic performance, involving local farmers expert teff growers in a participatory variety evaluation. Our analyses revealed environmental patterns of teff genetic diversity and allowed us to identify ten genetic clusters associated with climate variation and with uneven spatial distribution. A genome wide association study (GWAS) was used to identify loci and candidate genes related to phenology, yield, local adaptation, and farmers' appreciation. The estimated teff genomic offset under climate change scenarios highlighted an area around lake Tana where teff cropping may be most vulnerable to climate change. Our results show that transdisciplinary approaches may efficiently propel untapped NUS farmer varieties into modern breeding to foster more resilient and sustainable cropping systems.

Introduction

Large-scale, high-yielding cropping systems rely on a remarkable small set of crops. Approximately half of the global farming land is devoted to maize, wheat, rice, and soybean (FAOSTAT, 2021) and the overall composition of worldwide food systems is uniform (Khoury et al., 2014). Yet, even today, hundreds of neglected or underutilized species (NUS) are actively cultivated in highly diversified, small-scale cropping systems, where they support the livelihoods of millions of people (Jamnadass et al., 2020). NUS benefited of scant research and breeding improvement. Their diversity is not only a proxy of pedoclimatic diversification of cropping systems, but also reflects socioeconomic diversity and cultural heritage of local farmers (Tadele and Z, 2019). Rich NUS agrobiodiversity is still conserved *in situ* in smallholder agriculture systems, where the selection and cultivation conducted by local farmers resulted in the stabilization of farmer varieties largely untapped by breeding. A comprehensive, transdisciplinary characterization of NUS farmer varieties that takes into consideration diversity, adaptation, and farmer-consumer preferences may unlock the potential of NUS for the sustainable intensification of farming systems (Dawson et al., 2018).

The research community can now leverage the big data revolution to bridge the gap between NUS and 21st century agriculture (Dawson et al., 2019). Genomic tools enable a rapid and cost-effective characterization of large germplasm collections to unlock agrobiodiversity for breeding (Poland, 2015), identify genetic factors responsible for traits of agronomic interest (Mascher et al., 2019), and accelerate genetic gains (Juliana et al., 2019). Genomic data can be combined with climatic data to gain insights into locus-specific adaptation (Lasky et al., 2015) and to estimate genomic vulnerability under climate change scenarios (Aguirre-Liguori et al., 2021). Data-driven methods can also be applied to characterize the socioeconomic contexts in which crops are grown (Van Etten et al., 2019), generating insights that are critical to understand cropping dynamics in smallholder farming systems (Fan and Rue, 2020; Terlau et al., 2019). Participatory varietal selection (PVS) approaches directly involve farmers in the evaluation of breeding

materials (Ceccarelli and Grando, 2007). Previous studies shown that PVS can be combined with genomic data to identify genomic loci responsible for farmers' appreciation (Kidane et al., 2017) and model local crop performance in farmer fields (de Sousa et al., 2021).

The Ethiopian highlands are paradigmatic of challenging agricultural ecosystems where NUS farmer varieties are widely cultivated. In Ethiopia, 85 million people live in rural areas; most of them are subsistence-based smallholder farmers that are responsible for about 90% of the cultivated land and agricultural output of the country (Bachewe and Taffesse, 2018). Teff (*Eragrostis tef*) is an annual, self-pollinating, and allotetraploid grass of the *Chloridoideae* subfamily (Ketema, 1997). It is widely grown in Ethiopia as a staple crop, where it is valued for its nutritional and health benefits, resilience in marginal and semi-arid environments and cultural importance (D'Andrea, 2008; Ketema, 1997). Teff was likely domesticated in the northern Ethiopian Highlands from the allotetraploid *Eragrostis pilosa*, but the timing of its initial cultivation and the identity of its diploid ancestors remain poorly understood (D'Andrea, 2008; Ingram and Doyle, 2003; VanBuren et al., 2020). Nowadays, thousands of local landraces are grown in a wide range of agroecologies, displaying broad environmental adaptation and phenotypic diversity (Woldeyohannes et al., 2020). Breeding efforts on teff have been underway since decades, and segregant and mutagenized populations are available (Cannarozzi et al., 2018; Q Zhu et al., 2012), yet teff yields remain still much lower than potentially attainable and substantially lower than those of other cereals grown in the region (Girma et al., 2014; Woldeyohannes et al., 2020). The full breeding potential of teff farmer varieties is still undisclosed, but research in teff is rapidly evolving; a draft genome sequence (Cannarozzi et al., 2014) and a high quality genome sequence (VanBuren et al., 2020) recently brought this species to the international genomics research spotlight. A data-driven research approach may unlock the full potential of teff agrobiodiversity and propel breeding for new improved varieties matching local needs.

Here, we report a transdisciplinary data-driven approach to characterize NUS genetic, agronomic, and climatic diversity, using teff as a case study. We selected and genotyped 321 teff farmer varieties derived from landraces and 45 teff improved lines with high-density molecular markers, and we characterized their agronomic performance at two locations in Ethiopia. Experienced teff farmers (15 women and 20 men) were asked to evaluate the teff genotypes, providing qualitative information that could allow us to prioritize better adapted genetic materials. We derived current and projected climate data at the sampling locations of teff accessions and used them to estimate genetic offset under climate change scenarios. We combined phenotypic, bioclimatic and farmers' evaluations data in a genome-wide association study framework to identify genomic loci and candidate genes with relevance for adaptation, performance, and farmers' preferences. We discuss the potential of data-driven participatory approaches to characterize NUS diversity and to highlight their potential for the sustainable intensification of farming systems.

Results and discussion

Teff farmer varieties harness broad genetic diversity

We assembled a representative collection of teff cultivated in Ethiopia, hereafter named Ethiopian Teff Diversity Panel (EtDP). The EtDP comprises 321 farmer varieties spanning the entire geographical and agroecological range of teff cultivation in Ethiopia, from the sub-moist lowlands of Tigray in the North to the moist lowlands of Oromia in the South, and from the sub-humid lowlands of Benishangul and Gumuz in the West to the sub-humid mid-highlands of Oromia in the East (Figure 1 – figure supplement 1, Supplemental File 1A) (MoA, 2000). The EtDP also includes 45 improved varieties released since the first breeding efforts on teff and up to the moment of the EtDP creation. A selection of seven *Eragrostis spp.*, putative wild relatives of teff, was included as outgroup. The genomic diversity of the EtDP was assessed starting from 12,153 high-quality, genome wide

single nucleotide polymorphisms (SNPs) derived from double digest restriction-site associated sequencing (ddRAD-seq) of individual accessions, followed by filtering for variant call quality and linkage disequilibrium (LD) pruning. The EtDP genomes could be grouped in 1,240 haplotype blocks (Supplemental File 1B). Chromosomes showed consistently higher pericentromeric LD, with localized LD peaks in telomeric regions (Figure 1 – figure supplement 2). The A and B sub genomes showed comparable yet different LD profiles, possibly due to their specificity in terms of dominance, content of transposable elements, and overall limited homoeologous exchange (VanBuren et al., 2020).

Population structure analyses, performed on a LD-pruned set of SNP makers, showed that EtDP accessions displayed varying degrees of genetic admixture (Fig. 1A) and could be best summarized by ten DAPC clusters (Figure 1 – figure supplement 3). About 15% of the genetic variability in the teff panel could be explained by the first three principal components (PCs) of LD-pruned SNP data (Fig. 1B-C). Teff accessions from Tigray, in Northern Ethiopia, were markedly separated from the rest, and mainly belonged to cluster 7 (Figure 1D). Breeding lines showed a narrower genetic base compared to the diversity maintained by Ethiopian farmers; these materials could not represent the broad diversity available within the EtDP (Figure 1 – figure supplement 4), and predominantly belonged to clusters 2, 8 and 10 (Figure 3D-E). Genetic clusters are a proxy of teff landraces diversity and may support breeding efforts by improving the identification of parent lines for genomic selection to counter the depletion of allelic diversity (Heffner et al., 2009), or establishing breeding groups to explore heterotic potential in teff (Boeven et al., 2016). The preservation and valorization of gene pools could allow to respond to future needs and priorities in light of changing climates and dynamic consumer preferences (Araya et al., 2011; Woldeyohannes et al., 2020).

<u>The distribution of teff genetic variation is associated with geographic and environmental factors</u>

Landraces evolve at the interface of natural and anthropogenic selection (Casañas et al., 2017). Hence, we hypothesized that teff genetic clusters might be associated with local environmental conditions. The EtDP showed limited stratification, as evidenced by the high levels of admixture and low genetic variance explained by the PCA (Figure 1 – figure supplement 5; Figure 1 B-C). Yet, some genetic clusters could be associated to agroecological zones with different temperature and precipitation conditions (Figure 1 – figure supplement 6; Figure 2 A-B; Figure 2-Supplement 1). Cluster 4 and 7 came from hot sampling locations, while cluster and 6 came from colder and wetter areas (Figure 2 – figure supplement 1). Extant landraces diversity is not only contributed by climate, but also by seed circulation. Studies have shown that cultural and social dynamics, such as belonging to similar or different ethnolinguistic groups, are key factors in shaping seed exchange networks (Labeyrie et al., 2016; Samberg et al., 2013). In Ethiopia, regional districts are markers of cultural and historical diversity, and smallholder farmers have limited capacity for long-distance travel. Thus, we hypothesized that regional distinctions could impact seed exchanges and spatial patterns of teff genetic diversity.

The district of sampling of teff landraces in the EtDP was therefore used to aggregate accessions and calculate genetic distances as a measure of fixation index (F_{st}). Although our experimental design does not allow to fully untangle the effect of geography, environment, and social factors, we found a pattern of genetic variation distribution that can be associated to an isolation by distance and adaptation process, where F_{st} values are significantly associated with geographic distance (Mantel r = 0.31; p = 9e-04) and environmental distance (Mantel r =0.352; p =0.0137) (Fig. 2C-D). Accessions from East Tigray (*Misraquawi*) showed the highest separation from the collection, followed by East Oromia (*Misraq Harerge*) and West Amhara (*Agew Awi*) (Fig. 2E). Tigray is believed to be the center of teff domestication, as earlier reports identified there much of its diversity (Costanza et al., 1979) and archaeobotanical excavations reported its cultivation since the Pre Aksumite period (before I Century CE) (D'Andrea, 2008). However, limited

archeological information is available from other parts of the country, which may disclose other sites of early domestication of teff. When integrating the putative teff wild relatives *Eragrostis curvula* and *Eragrostis pilosa* in our teff landraces' phylogeny, we found that they do not group with any specific DAPC cluster (Figure 2 – figure supplement 2).

Participatory evaluation of the teff diversity prioritizes genetic materials for breeding

The EtDP was phenotyped for agronomic and farmer appreciation traits during the main cropping season in two locations in Ethiopia (Fig. 1D), representing areas with high potential for teff cultivation. Experienced teff farmers including 15 women and 20 men provided their evaluations on overall appreciation (OA) and panicle appreciation (PA) across the EtDP. Both women and men farmers were selected because they were teff growers, and their responses were analyzed separately to highlight distinct criteria for evaluating traits. Previous studies conducted in Ethiopia showed that agricultural traits relevant to marketability, food and drink preparations, animal feed and construction are all perceived differently by women and men farmers (Assefa et al., 2014; Mancini et al., 2017). Regardless of gender, farmer evaluations were highly heritable and fully comparable to agronomic traits commonly targeted by breeding. OA provided by farmers across genders and across locations had a broad sense heritability (H^2) of 0.81 (Table 1). H^2 of grain yield combined across the same two locations was 0.42 (Table 2).

The high H^2 achieved by farmers' OA may be because farmers, in providing their overall evaluation, not only consider yield but also yield components with high heritability (Table 3). Farmers' appreciation of teff genotypes was strongly associated with yield ($p < 0.001$) and its components regardless of gender, most notably plant height ($p < 0.001$) and grain filling period ($p < 0.05$) (Fig. 3A, Figure3–figure supplement 1, Table 3). Plant height and grain filling period were very important for men, while women OA was highly correlated with biomass yield and

panicle weight (Table 3). Different dynamics of trait prioritization by men and women are expected, and may reflect gender roles in the value chain (Weltzien et al., 2019): while men are mostly endowed with field work, women are concerned with marketability of grains (Mancini et al., 2017). However, men and women alike were expert teff growers, as reflected by the consistency of their evaluations (Fig. 3). Previous studies showed that farmers' appreciation is genetically determined, and may be used to perform genomic prediction (de Sousa et al., 2021) and identify genomic associations (Kidane et al., 2017).

The top ranking teff accessions according to men and women farmers captured different genetic backgrounds (Fig. 3B), but the same trait combinations (Fig. 3C). This suggested that farmers were consistently preferring the same teff ideotype regardless of their gender and regardless of the genetic background and geographic provenance of the accession. High yielding, high-biomass and fast maturing varieties were preferred. Individually, teff improved varieties showed high OA and high panicle length, which is a main component of OA (Fig. 3D-E). This is supported by the fact that the evaluation fields were high-potential areas for teff cultivation. Genetic cluster 2, that is representative of most breeding materials, is associated with longer days to heading and maturity, higher plant heights and panicle lengths, greater number of total tillers and higher yields (Figure 3 – figure supplement 2). However, several farmer varieties from different genetic backgrounds recorded comparable, at times superior, performance than improved lines (Fig. 3G). By selecting farmer varieties that outperform improved varieties' performances in target traits for breeding, it may be possible to prioritize landrace accessions for teff improvement (Fig. 3D-E) or even immediately make them available to farmers, as suggested by previous experiences in wheat (Fadda et al., 2020). Short maturation time is paramount to achieve harvestable yield in areas exposed to terminal drought (Mengistu and Mekonnen, 2012), and it is expectedly a major component of farmers' OA (Fig. 3A). The time of maturation is therefore an obvious target trait for teff breeding, although challenging to be combined with higher attainable yield. We found that many farmer varieties had a shorter grain

filling period than most improved lines, and that genotype EBI *9551* combined this trait with higher yield and farmers' appreciation (Fig. 3G).

Participatory, climatic, and agronomic information to support teff breeding

The data deriving from the transdisciplinary characterization of the EtDP was integrated in a GWAS framework to identify genomic loci associated with agronomic performance, local adaptation, and farmer preferences. These loci could be then used in marker assisted selection, genomic selection, or new breeding technologies to support teff improvement. The GWAS led to the identification of a total of 193 unique quantitative trait nucleotides (QTNs) (Supplemental File 1C). Of these, 101 QTNs were associated with bioclimatic variables at sampling sites. Seventeen more QTN were associated with farmers' appreciation. A total of 115 QTNs could be grouped in 59 haplotype blocks marking the co-inheritance of contiguous SNPs. We focused on these LD blocks to identify homology of predicted teff proteins with protein sequences of *Arabidopsis thaliana* and *Zea mays* (Supplemental File 1D). Among others, we identified a locus on chromosome 2A that was associated with grain yield, grain filling rate and overall appreciation (lcl|2A-14415768) (Figure 5; Figure 5 – figure supplement 2). The LD region targeted by this QTN harbors 39 gene models including two 60s ribosomal subunits and several homologs of maize genes with suggestive function in relation to yield determination. *Et_2A_015515*, in this block, is a homolog of a maize Serine/threonine protein kinase 3, belonging to abroad class of proteins that was associated with inflorescence development (McSteen et al., 2007) and grain yield (Jia et al., 2020) in maize. These findings, although preliminary, may support the development of markers for teff improvement as well as directing local and international research towards loci of particular relevance for teff improvement. We further explored the EtDP data using a gradient forest (GF) machine learning algorithm to calculate the contribution of environmental gradients to genomic variation in teff, and to estimate its genomic offset, or vulnerability, under projected climates. We found that the turnover of allele frequencies across the cropping area

was best predicted by geographic variables (Moran's eigenvector map variables, MEM) and by precipitation indicators, particularly precipitation of the coldest quarter (bio19) and precipitation of the wettest month (bio13) (Figure 4 – figure supplement 1). The GF allowed us to model climate-driven genomic variation patterns across the landscape (Fig. 4A-B). Approximately a quarter of the SNPs (3,049 in 521 LD blocks) were predictive of the GF model: of these, 176 showed F_{st} values in the 99th percentile of the distribution (Supplemental File 1E).

The GF outcome was then linked to GWAS to identify genomic loci associated to climate, agronomic performance, farmers' preference, and adaptive potential of teff (Figure 4 – figure supplement 2). We found that LD blocks containing QTNs for days to maturity were strongly associated with the GF model (Figure 4 – figure supplement 3), supporting the importance of phenology in teff adaptation and geographic distribution (Woldeyohannes et al., 2020). On chromosome 1A at 32.3 Mb, three QTNs for days to maturity colocalized with a large significance peak for precipitation of driest month (bio14) and PC2 of bioclimatic variables, representing seasonality (Figure 4 – figure supplement 2; Figure 5 – figure supplement 1; Figure 5 – figure supplement 2). This LD block harbors 176 gene models, among which four with predicted proteins with high homology with maize and Arabidopsis proteins (Supplemental File 1D). The product of *Et_1A_007229* is predicted to be homologous to a phosphatidylinositol kinase that is involved in flower development and has been shown to influence floral transition in condition of abiotic stress (Akhter et al., 2016). Days to maturity were also associated with a LD block at 20.8 Mb on chromosome 6A that was predictive for the GF model (Figure 4 – figure supplement 2; Figure 4 – figure supplement 3). In this block, nine gene products share homology with Arabidopsis and maize (Supplemental File 1D). The protein encoded by *Et_6A_046800* is homologous to an ETO1-like protein involved in the regulation of ethylene synthesis in Arabidopsis (Wang et al., 2004). Ethylene is a key plant hormone that has been shown to be related to spike development and senescence (Valluru et al., 2017). These candidate genes have not been yet validated. With an increasing refinement of teff gene annotations, corroborated by

reverse genetic approaches (Qihui Zhu et al., 2012), it will be possible to validate genes underlying traits of interest. Teff breeding could then fully benefit from targeted editing (Lemmon et al., 2018) to speed up the development of new varieties with improved yield, local adaptation and adherence to local preferences.

The genome-wide teff adaptive potential across the landscape varied in magnitude and distribution according to different predicted climate scenarios for 2070 (Figure 4 – figure supplement 4). We used this data to compute the genomic-adaptive offset between current and future climate scenarios to identify vulnerable areas (Fig. 4C, Figure 4 – figure supplement 5). In all representative concentration pathways (RCPs), the highest offset was predicted in the north-western highlands of the Amhara region, south of lake Tana (Fig. 4C, Figure 4 – figure supplement 4). Compared to other regions of the country, we found a decreasing trend of rainfall change in this region across all emission scenarios (Figure 4 – figure supplement 5; Figure 4 – figure supplement 6). In this area hot nights are projected to increase more quickly than hot days, with the most marked increases expected to be experienced in the July, August, September season (Figure 4 – figure supplement 7). Decreasing trends of rainfall during the main growing season are predicted in all projected scenarios, suggesting that seasonality might critically teff development stages (Figure 4 – figure supplement 6). In light of these results, a valid adaptation strategy could be the assisted migration of teff genotypes from areas of higher vulnerability to areas of lower vulnerability (Rhoné et al., 2020). However, crop migration and varietal replacement strategies will need to take into account ecological and socio-economic factors, including the impacts on existing ecosystems and on farmers' adoption of migrated varieties (Sloat et al., 2020).

Conclusion

A comprehensive interpretation of crop performance is key to a sustainable intensification that embraces cultural and agricultural diversity of cropping systems. While significant successes and even a plateau might have been reached

in optimal growing environments where most common crops are cultivated, there is ample opportunity to enhance productivity in marginal growing environments (Godfray et al., 2010). The success of crop varieties is not only determined by yield performance, but also by adaptation to local environments and cultural needs (Weltzien et al., 2019). The integration of genomic, climatic, and phenotyping diversity in a participatory framework may help tailor varietal development for local adaptation. The involvement of farmers in varietal evaluation is increasingly utilized in a quantitative framework to guide breeding choices in combination with genomic data (Annicchiarico et al., 2019; de Sousa et al., 2021).

Transdisciplinary methods may support the integration of smallholder farmers in modern breeding and agricultural value chains. Modern data-driven research may then efficiently harness the genetic diversity generated in farmers' fields. Teff is rapidly emerging from the NUS status and is projected towards modern breeding. Genebank genomics may be systematically used to map its large agrobiodiversity (Woldeyohannes et al., 2020), fully disclosing the opportunity to use next generation breeding technologies (Varshney et al., 2021) and large-scale genomic selection (Poland, 2015). Multi-location trials are needed to capture the range of genotype by environment (G x E) interactions that influence agronomic performance of teff, which we could only partially characterize. Decentralized varietal evaluation approaches can be used to scale up the testing of teff genetic resources (de Sousa et al., 2021; van Etten et al., 2019) to better capture G x E, produce new varieties with higher local adaptation, and resulting in higher farmers' varietal adoption. An extensive socioeconomic characterization of teff cropping, including extensive farmers interviews (Labeyrie et al., 2016), may further unveil the dynamics of teff seed exchange and associated flow of information and knowledge related to local cropping (Occelli et al., 2021). Public sector breeding beyond NUS may further enhance the combination of data-driven research with participatory approaches to improve customer and product profiling to achieve social impact as cost effective as possible.

The IPCC reports indicate that East Africa will experience an increase in aridity and agricultural droughts, with a substantially higher frequency of 'hot' days and nights (IPCC, 2017). Temperature increases are also expected to result in more intense heat waves and higher evapotranspiration rates, which, coupled with the altered rainfall patterns, will affect agricultural productivity. Enhancing NUS farmer varieties offers promising opportunities to tackle food insecurity resulting from climate change in smallholder farming settings and beyond. NUS have enormous untapped potential for improvement that is hampered by lack of tools and knowledge (Yerima and Achigan-Dako, 2021), but our analyses show that their characterization is at hand. As NUS proceed towards mainstream breeding, the collaborative effort of scientists, breeders, and farmers will unlock their full potential for sustainable intensification of farming systems.

Materials and Methods

Plant materials

The *E. teff* diversity panel (EtDP) used in this study was derived from a larger teff collection of 3,850 accessions held at the Ethiopian Biodiversity Institute (EBI) (Addis Abeba, Ethiopia), which represents the world's largest active teff collection and was amplified and characterized by Woldeyohannes and collaborators (Woldeyohannes et al., 2020). Criteria for selecting the EtDP from the larger EBI collection were the following: i) visible morphological variation for panicle traits, ii) geographical and agroecological representativeness, iii) apparent grain yield potential, iv) presence of different maturity groups, and v) presence of associated traditional names of specific landraces (*e.g. Bunninye*, *Murri*, *Fesho*). The EtDP includes 321 farmer varieties, sided by all 45 *E. teff* improved varieties released since the beginning of teff breeding program and until the selection of the EtDP. Improved varieties were obtained by a selection conducted by Ethiopian agricultural research centers. Seven accessions of teff wild relatives *E. pilosa* and *E. curvula* were also included in the collection (Supplemental File 1A). Landraces

were purified selecting and reproducing one single panicle representative of each accession. The ITPGRFA defines a crop variety as "[…] defined by the reproducible expression of its distinguishing and other genetic characteristics" (Ho, 2011). For the scope of this paper, we define farmer varieties as uniform genotypes derived from the purification of *ex-situ* accessions (*i.e.* landraces) collected from local farmers. Farmer varieties are therefore a proxy of landraces originally collected in farmer fields and are discussed as such.

Sequencing and variant calling

Seeds of the EtDP were germinated in pots at the EBI in 2018, and at least three seedlings were harvested and pooled per accession. Genomic DNA was extracted from pooled seedlings at the EBI laboratories using the GenElute Plant Genomic DNA Miniprep Kit (Sigma-Aldrich, St. Louis, MO, USA) following the manufacturer's instructions. DNA quality was checked by electrophoresis on 1% agarose gel and using a NanoDrop ND-1000 spectrophotometer and sent to IGATech (Udine, Italy) for sequencing. Genomic libraries were produced using *SphI* and *MboI* restriction enzymes in a custom protocol for the production of double digestion restriction site-associated DNA markers (ddRAD) (Peterson et al., 2012). In short, ddRAD is based on a double restriction of the target DNA, typically using a rare cutter and a frequent cutter, followed by sequencing of restriction fragments to reduce the complexity of the target genome. ddRAD libraries were sequenced with V4 chemistry on Illumina HiSeq2500 sequencer (Illumina, San Diego, CA) with 125 cycles in a paired-end mode. Reads were demultiplexed using the *process_radtags* utility included in Stacks v2.0 (Catchen et al., 2013) and analyzed for quality control with the FastQC tool (v.0.11.5). High-quality paired-end reads of each individual were mapped against the *Eragrostis tef* reference genome (version 3, available from CoGe under ID 50954) (VanBuren et al., 2020) with BWA (Burrows-Weeler-Aligner v.0.7.12) using the MEM algorithm with standard parameters (Li, 2013). Alignments were sorted and indexed with PicardTools (http://broadinstitute.github.io/picard/) and samtools (Li et al., 2009).

Single nucleotide variants were identified with GATK (McKenna et al., 2010) HaplotypeCaller algorithm (version 4.2.0), run in per-sample mode followed by a joint genotyping step completed by GenotypeVCFs tool. Raw variants were filtered out using the VariantFiltration and SelectVariants GATK functions with the following criteria: monomorphic or multiallelic sites, QUAL < 30; QD < 2.0; MQ < 40.0; AF < 0.01; DP < 580; *SNP clusters* defined as three or more variants located within windows of 5 bp. For each accession, SNPs with a total read count of <3 were set to NA. Variants were discarded if located on unanchored contigs, InDels, missing data >20%, heterozygosity >15%.

Bioclimatic characterization

GPS coordinates of EtDP teff landraces were derived from EBI passport data and projected onto the map of Ethiopia using R/raster (Hijmans and van Etten, 2012). Teff landraces were projected onto the agroecological-zones map of Ethiopia provided by the Ethiopian Institute of Agricultural Research (EIAR) (MoA, 2000), which subdivides Ethiopia into different zones according to altitudinal ranges and temperature and rainfall patterns. Altitudes were assigned to each landrace based on GPS coordinates, using the CGIAR SRTM database at 90 m resolution (Reuter et al., 2007). Current climate data (1970-2000 averages) relative to teff landraces' sampling sites were retrieved from the WorldClim 2 database of global interpolated climate data (Fick and Hijmans, 2017) at the highest available spatial resolution of 2.5' (approximately 4 x 4 km). Collinearity among historical bioclimatic variables was previously checked with the *ensemble.VIF()* function in R/BiodiversityR (Kindt and Coe, 2005). Only variables with a variation inflation factor (VIF) below 10 were retained, namely bio2, bio3, bio4, bio9, bio13, bio14, bio15, bio18 and bio19. The Hadley Centre Global Environmental Model 2-Earth System (HadGEM2-ES) (Jones et al., 2011) under the fifth phase of the Coupled Model Intercomparison Project (CMIP5) protocols simulations was used to retrieve future climate scenarios at the following representative concentration pathways (RCPs): RCP 2.6, RCP 4.5, RCP 6.0 and RCP 8.5.

Phenotyping and participatory variety selection

The EtDP was phenotyped in common garden experiments in two high potential teff growing locations in Ethiopia, Adet (Amhara, 11° 16′32″ N, 37° 29′30″ E) and Akaki (Oromia, 8°50'07.6"N, 38°49'58.3"E), under rainfed conditions during the main cropping season of 2018 (July-November). Adet is the main experimental site of the Amhara Regional Agricultural Research Institute, at an altitude of 2,240 meter above sea level. Soil type is vertisol, climate is moist cool, and average annual rainfall is 1,250 mm. Akaki is a sub site of the Debre Zeit Agricultural Research Center, with an altitude of 2,200 meter above sea level. Also in this case, soil type is vertisol, climate is moist cool and average annual rainfall is 1,055 mm. Accessions were planted in two replications per site using an alpha lattice design, in plots consisting of three rows of 1 m in length and 0.2 m inter-row distance. Three phenological traits, days to 50% heading (DH), days to 90% maturity (DM), and grain filling period (GFP) were recorded on whole plots in each environment. Morphology and agronomic traits were recorded from five randomly selected teff plants per plot: plant height (PH, cm), panicle length (PL, in cm), number of primary branches per main shoot panicle (PBPM), number of total tillers (TT), first culm length (CLF, in cm), first culm internode diameter (CDF, in mm), panicle weight (PW, in grams), panicle yield (PY, in grams), grain yield (GY, ton/ha), biomass yield (BY, ton/ha), harvest index (HI), grain yield filling rate (GFR, kg/ha/day), and biomass production rate (BPR, kg/ha/days). Qualitative data was sourced from the characterization performed by Woldeyohannes and collaborators (Woldeyohannes et al., 2020).

A participatory variety selection (PVS) was conducted in the two locations, involving 35 experienced teff farmers: 15 men and 10 women in Adet, five men and five women in Akaki. PVS was conducted close to physiological maturity in each location so to maximize variation between plots. Participating farmers were selected with the help of agricultural officers. They all had experience on teff production and were recognized as a local agricultural expert in teff cultivation.

Being local farmers, they spoke different languages (Amharic and Oromo) and local interpreters were employed. Farmers were engaged in focus group discussions prior to the PVS to discuss most relevant traits in teff cultivation and to attend training on how to perform the PVS. During the evaluation, farmers were divided into gender-homogeneous groups with five people each. Groups were conducted across the field from random entry points and asked to evaluate two traits: panicle appreciation (PA) and overall appreciation (OA). PVS traits were given on a Likert scale from 1 (poor) to 5 (excellent), in a way answering a question in the form of: “how much do you like the [panicle appearance / overall appearance] of this plot from one to five?”. Farmers provided their scores simultaneously so that within-group scoring bias was reduced. Each farmers’ score was recorded individually. We used Plackett-Luce model (Turner et al., 2020) to analyze farmers’ OA. Agronomic traits from tested genotypes were linked to farmer’s overall appreciation with a linear model using an Alternating Directions Method of Multipliers (ADMM) algorithm proposed by (Yıldız et al., 2020).

<u>Phenotypic data analyses</u>

Best linear unbiased predictions (BLUP) of agronomic and PVS traits were computed with R/ASReml (Gilmour et al., 2014). BLUPs for agronomic traits were derived from the general model in Eq. (S1):

$$y_{ijkn} = \mu + g_i + l_k + gl_{ik} + e \qquad \text{Eq. (S1)}$$

Where the observed phenotypic value is y_{ik}, $\mu$is the overall mean of the population, $g_i$is the random effect for the *i*th genotype *g*, l_k is the fixed effect for the *k*th location, $gl_{ik}$is the random effect interaction between genotype and location, and $e$is the error. For calculation of BLUPs with a single location, the data was sub-set by location and the model in Eq. (S1) was simplified accordingly. Broad-sense heritability (H^2) of agronomic traits was derived from the variance component estimates deriving from Eq. (S1) as follows:

$$H^2 = \frac{\sigma_g}{\left(\sigma_g + \frac{\sigma_{gl}}{n_{loc}} + \frac{\sigma_e}{n_{rep} * n_{loc}}\right)} \quad \text{Eq. (S2)}$$

In Eq. (2), σ_g is the variance component of genotypes,σ_{gl} is the genotype by location variance, and σ_e is the error variance. $n_{loc} n_{rep}$are the number of locations and replications, respectively. For calculation of *H²* within locations (*i.e.* repeatability), Eq. (S2) was simplified accordingly.

The derivation of PVS BLUPs and *H²* was like that used for agronomic traits except for the fact that gender of farmers was considered. BLUPs for PVS were obtained from the model in Eq. (S3):

$$y_{ikm} = \mu + g_i + l_k + p_m + gl_{ik} + gp_{im} + pl_{mk} + e \quad \text{Eq. (S3)}$$

Where y_{ikm} is the observed PVS score, and μ, g_i, l_k, and gl_{ik} are as in Eq. (S1) and p_m is the random effect for farmer gender. Accordingly, $gp_{im}$is the random effect of the interaction between genotype and gender andpl_{mk} is the random interaction between gender *m* and the *k^th^* location. For calculation of BLUPs specific for gender, location and gender by locations, Eq. (S3) was simplified accordingly. *H²*for PVS traits was derived from the following formula:

$$H^2 = \frac{\sigma_g}{\left(\sigma_g + \frac{\sigma_{gl}}{n_{loc}} + \frac{\sigma_{gm}}{n_{gender}} + \frac{\sigma_e}{n_{rep} * n_{loc} * n_{gender} * n_{farmer}}\right)} \quad \text{Eq. (S4)}$$

In Eq. (S4), σ_g is the variance component of genotypes, σ_{gl} is the genotype by location variance, $\sigma_{gm}$is the genotype by gender variance, and σ_e is the error variance. n_{loc}, n_{gender}, and $n_{rep}$are the number of locations, genders and replications, respectively. For calculation of *H²* (*i.e.* repeatability) by gender and by location, Eq. (S4) was simplified accordingly. The 90th percentile of the OA distribution was considered to identify top ranking accessions for men and women.

Farmer varieties were benchmarked with the fourth quartile of the distribution of improved lines for all scored traits.

Genetic diversity

Phylogenetic relationships in the EtDP were assessed on a pruned set of SNP markers with MAF > 0.05. Pruning was performed with the PLINK (Purcell et al., 2007) indep-pairwise function on a 100 SNPs window moving in 10 SNP steps with a linkage disequilibrium (LD) r^2 threshold of 0.3. Pairwise identity by descent (IBS) was calculated by PLINK and visualized with custom scripts in R (R Core Team, 2018). A principal component analysis (PCA) and a discriminant analysis of principal components (DAPC) were performed with R/adegenet (Jombart, 2008). The optimal number of clusters (K) for the DAPC was identified using adegenet::find.cluster(). Bayesian Information Criterion (BIC) statistics were computed at increasing values of K to measure the goodness of fit at each K using 365 PCs and default settings. Admixture (Alexander and Lange, 2011) was run testing 2 to 25 K clusters using the default termination criterion. Each iteration was run using different random seeds, and parameter standard errors were estimated using 2,000 bootstrap replicates. A 5-fold cross-validation procedure was used to identify the most likely value of K. The correlation of the residual difference between the true genotypes and the genotypes predicted by the model was estimated using EvalAdmix (Garcia-Erill and Albrechtsen, 2020). A neighbor joining (NJ) tree was developed computing genetic distances using the Tajima-Nei method (Tajima and Nei, 1984) and performing 500 bootstrap resampling, using MEGA X (Kumar et al., 2018). Different NJ tree visualizations were produced using R/ggtree (Yu et al., 2017). Genotypic data of putative teff wild relatives *Eragrostis curvula* and *Eragrostis Pilosa* was integrated in the NJ phylogeny. A set of putative SNPs shared between wild relatives and cultivated teff was derived as described for the EtDP.

LD analyses were performed on SNPs with MAF > 0.05. Average pairwise r^2 for all markers within a window of ±5 Mb was estimated using R/Ldheatmap (Shin et

al., 2006). The LD was plotted against physical positions, averaging pairwise r^2 values for each chromosome over sliding window considering portions equal to 5% of each chromosome's physical length. LD decay was then estimated for each of the chromosomes according to the Hill and Weir (Hill and Weir, 1988) equation using a threshold of $r^2 = 0.3$. Haplotype blocks were estimated using the PLINK --blocks function with default settings and following the interpretation of Gabriel and colleagues (Gabriel et al., 2002).

Climatic diversity and gradient forest

Agroecological and bioclimatic variation analyses were performed on georeferenced materials of the EtDP. The distribution of the DAPC genetic clusters across agroecological zones was mapped via R/raster. After aggregating teff georeferenced accessions in Ethiopian administrative regions at the second level (districts), pairwise F_{st} (Weir and Cockerham, 1984) was calculated across all SNP markers for all areas accounting at least 5 individuals. Centroid coordinates of the accessions within each district were used to estimate geographic distances, while environmental distances were calculated by averaging the value of non-correlated historical bioclimatic variables and altitude. A measure of environmental distance between each accession was thus calculated as pairwise Euclidean differences between locations. A Mantel test with a Monte Carlo method (9,999 replications) was implemented in R/ade4 (Dray and Dufour, 2007) to check associations between linearized F_{st} ($F_{st}/1-F_{st}$) and geographic and environmental distances.

The teff cropping area was defined by the union of all polygons representing agroecological zones in which at least two teff landraces were sampled. Significant associations between genetic clusters and agroecological zones and administrative regions were assessed using Pearson's Chi-squared test of independence. Pairwise Wilcoxon rank sum test was used to test the significance ($p < 0.05$) of differences in bioclimatic variables among DAPC clusters. A gradient forest (GF) machine-learning approach implemented in R/gradientForest (Ellis et al., 2012; Fitzpatrick and Keller, 2015), was used to map the turnover in allele

frequencies using non-linear functions of environmental gradients with historical and projected climates. The GF was developed using historical non-collinear bioclimatic variables and Moran's Eigenvector Map (MEM) variables representing climatic and geographic diversity in the sample, respectively. MEM variables were derived from geographic coordinates at sampling locations of the landraces in the EtDP (Dray et al., 2006; Griffith and Peres-Neto, 2006) and were calculated with *dbmem()* in R/adespatial (Stéphane Dray et al., 2021). A function was built for each response variable (SNPs) using 500 regression trees. An aggregate function was created for all SNPs, whereas the bioclimatic variables and MEMs were used as predictors. The model was then run to predict teff genetic-geographic-climatic distribution on the teff cultivation range in Ethiopia. The GF model was also run using and projected climate data under different RCP scenarios.

Climate projections for areas of interest were analyzed to assess trends in rainfall and temperature. The twelve models best performing in the East Africa region according to IPCC (IPCC, 2017) were used to develop and ensemble projection of rainfall and temperature indices with Climate Data Operators (CDO) (Schulzweida, 2017) and custom R scripts. Projected data was compared with historical data to derive indices change in the interannual variability for the regions of interest.

Genome-Wide Association Studies

Quantitative trait nucleotides (QTNs) were mapped in a genome wide association study (GWAS). GWAS was performed with R/rMVP (Yin et al., 2020) using the Fixed and random model Circulating Probability Unification (FarmCPU) method (Liu et al., 2016) that incorporates corrections for population cryptic relatedness (Kinship). The first 10 genetic PCs were used as covariates. The Kinship was estimated using the method implemented by VanRaden (VanRaden, 2008). Both kinship and PCA were calculated using the subset of LD-pruned markers used for population genetics analysis. GWAS was run on bioclimatic variables, agronomic traits, and PVS traits. QTN were called when association surpassed a multiple testing correction with False Discovery Rate (FDR) of 5% using the R/q-value

(Storey et al., 2021). QTNs were assigned to the previously defined haplotype blocks. Blocks were extended by the chromosome-specific LD decay distance upstream and downstream and used as windows to search for candidate genes. The LD blocks thus obtained were combined with Fst and GF results to identify intersections across methods.

Teff gene annotations were retrieved from CoGe under id50954 (VanBuren et al., 2020). Nucleotide sequences of putative candidate genes were translated into the corresponding proteins and used as queries against Araport11 (Cheng et al., 2017) and the Maize reference proteome, available from UniProt (https://www.uniprot.org/) under the ID UP000007305. *E-value* of 10^{-20} and percentage of identity of 50% were used as threshold to retain blast hits on Arabidopsis and maize.

Acknowledgments

We wish to thank the men and women farmers who took part to the research, evaluating hundreds of genetic materials with great enthusiasm and competence in Adet and Akaki. These are their names in alphabetical order: Abebech, Adisse Abera, Adugawu Kere, Alebachew, Angoachi Fente, Askale, Ayana, Belachew Adimasu, Belay, Birtukan Ayele, Birhanu, Genanaw Dejene, Getasil, Habtamu Belay, Mame Seyum, Mamo, Misganaw Dagnaw, Regassa, Sefiwu Kassahun, Shashe, Tenagne Wubet, Tesfa Kefyalew, Tihune Dires, Tirusew, Tsehay Desie, Workinesh, Workineh Tsega, Wubet, Wubitu, Yechale Ayalneh, Yekoye, Yeniguse Worku, Yeniguse Wuletaw, Zebu Kebede, and Zina Yitay. We thank the Ethiopian Biodiversity Institute (EBI) for providing seed material and laboratory space to perform the DNA extraction in Ethiopia. Many thanks to Dr. Eleni Shiferaw and Dr. Basazen Fantahun Lakew for coordinating local support at EBI. Funding was provided by the Doctoral School in Agrobiodiversity at Scuola Superiore Sant'Anna.

Author Contributions

MEP, MD, conceived and supervised the research. CF and EAD contributed to the research design and data interpretation. ABW collected data, coordinated PVS and phenotyping, and conducted phenotypic analysis. MM conducted bioinformatics analysis with the contribution of SDI. LC conducted climatic data analysis with the contribution of JSA. ABW, SDI, LC, MM, KdS, and MD analyzed and interpreted data. SDI, MD, LC, MM drafted the manuscript and produced figures. All authors approved the final version of the manuscript.

Competing Interest Statement

The authors declare no conflict of interest.

Data availability

Teff accessions are available upon request from the Ethiopian Biodiversity Institute (EBI, http://www.ebi.gov.et/). Raw DNA sequencing reads are available on the Short Read Archive (https://www.ncbi.nlm.nih.gov/sra/) under BioProject accession number PRJNA758057. All scripts used for data analysis are available on GitHub at https://github.com/mdellh2o/TeffDiversityPanel.

References

Aguirre-Liguori JA, Ramírez-Barahona S, Gaut BS. 2021. The evolutionary genomics of species' responses to climate change. *Nat Ecol Evol 2021* 1–11. doi:10.1038/s41559-021-01526-9

Akhter S, Uddin MN, Jeong IS, Kim DW, Liu XM, Bahk JD. 2016. Role of Arabidopsis AtPI4Kγ3, a type II phosphoinositide 4-kinase, in abiotic stress responses and floral transition. *Plant Biotechnol J* **14**:215–230. doi:10.1111/pbi.12376

Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**:246. doi:10.1186/1471-2105-12-246

Annicchiarico P, Russi L, Romani M, Pecetti L, Nazzicari N. 2019. Farmer-participatory vs. conventional market-oriented breeding of inbred crops using phenotypic and genome-enabled approaches: A pea case study. *F Crop Res* **232**:30–39. doi:10.1016/j.fcr.2018.11.001

Araya A, Stroosnijder L, Girmay G, Keesstra SD. 2011. Crop coefficient, yield response to water stress and water productivity of teff (Eragrostis tef (Zucc.). *Agric Water Manag* **98**:775–783. doi:10.1016/j.agwat.2010.12.001

Assefa T, Sperling L, Dagne B, Argaw W, Tessema D, Beebe S. 2014. Participatory Plant Breeding with Traders and Farmers for White Pea Bean in Ethiopia. *J Agric Educ Ext* **20**:497–512. doi:10.1080/1389224X.2013.824385

Bachewe FN, Taffesse AS. 2018. Supply response of smallholder households in EthiopiaThe Economics of Teff: Exploring Ethiopia's Biggest Cash Crop. Washington, DC: International Food Policy Research Institute (IFPRI). pp. 181–204. doi:10.2499/9780896292833_08

Boeven PHG, Longin CFH, Würschum T. 2016. A unified framework for hybrid breeding and the establishment of heterotic groups in wheat. *Theor Appl Genet 2016 1296* **129**:1231–1245. doi:10.1007/S00122-016-2699-X

Cannarozzi G, Chanyalew S, Assefa K, Bekele A, Blösch R, Weichert A, Klauser

D, Plaza-Wüthrich S, Esfeld K, Jöst M, Rindisbacher A, Jifar H, Johnson-Chadwick V, Abate E, Wang W, Kamies R, Husein N, Kebede W, Tolosa K, Genet Y, Gebremeskel K, Ferede B, Mekbib F, Martinelli F, Pedersen HC, Rafudeen S, Hussein S, Tamiru M, Nakayama N, Robinson M, Barker I, Zeeman S, Tadele Z. 2018. Technology generation to dissemination: lessons learned from the tef improvement project. *Euphytica* **214**. doi:10.1007/s10681-018-2115-5

Cannarozzi G, Plaza-Wüthrich S, Esfeld K, Larti S, Wilson YS, Girma D, de Castro E, Chanyalew S, Blösch R, Farinelli L, Lyons E, Schneider M, Falquet L, Kuhlemeier C, Assefa K, Tadele Z. 2014. Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (Eragrostis tef). *BMC Genomics* **15**. doi:10.1186/1471-2164-15-581

Casañas F, Simó J, Casals J, Prohens J. 2017. Toward an Evolved Concept of Landrace. *Front Plant Sci* **0**:145. doi:10.3389/FPLS.2017.00145

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: An analysis tool set for population genomics. *Mol Ecol* **22**:3124–3140. doi:10.1111/mec.12354

Ceccarelli S, Grando S. 2007. Decentralized-participatory plant breeding: an example of demand driven research. *Euphytica* **155**:349–360. doi:10.1007/s10681-006-9336-8

Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD. 2017. Araport11: a complete reannotation of the {Arabidopsis} thaliana reference genome. *Plant J Cell Mol Biol* **89**:789–804. doi:10.1111/tpj.13415

Costanza SH, Dewet JMJJ, Harlan JR. 1979. Literature review and numerical taxonomy of Eragrostis tef (T'ef). *Econ Bot* **33**:413–424. doi:10.1007/BF02858337

D'Andrea AC. 2008. T'ef (Eragrostis tef) in ancient agricultural systems of highland Ethiopia. *Econ Bot* **62**:547–566. doi:10.1007/s12231-008-9053-4

Dawson IK, Hendre P, Powell W, Sila D, ... 2018. Supporting human nutrition in Africa through the integration of new and orphan crops into food systems.

World Agrofor …. doi:10.5716/WP18003.PDF

Dawson IK, Powell W, Hendre P, Bančič J, Hickey JM, Kindt R, Hoad S, Hale I, Jamnadass R. 2019. The role of genetics in mainstreaming the production of new and orphan crops to diversify food systems and support human nutrition. *New Phytol* **224**:37–54. doi:https://doi.org/10.1111/nph.15895

de Sousa K, van Etten J, Poland J, Fadda C, Jannink J-L, Kidane YG, Fantahun B, Mengistu DK, Pè ME, Solberg SO, Dell'Acqua M. 2021. Data-driven decentralized breeding increases prediction accuracy in a challenging crop production environment. *Commun Biol*. doi:10.1038/s42003-021-02463-w

Dray S, Dufour A-B. 2007. The ade4 Package: Implementing the Duality Diagram for Ecologists. *J Stat Softw* **22**:1–20. doi:10.18637/JSS.V022.I04

Dray S, Legendre P, Peres-Neto PR. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices ({PCNM}). *Ecol Modell* **196**:483–493. doi:10.1016/j.ecolmodel.2006.02.015

Ellis N, Smith SJ, Pitcher CR. 2012. Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93**:156–168. doi:https://doi.org/10.1890/11-0252.1

Fadda C, Mengistu DK, Kidane YG, Dell'Acqua M, Pè ME, Van Etten J. 2020. Integrating Conventional and Participatory Crop Improvement for Smallholder Agriculture Using the Seeds for Needs Approach: A Review. *Front Plant Sci*. doi:10.3389/fpls.2020.559515

Fan S, Rue C. 2020. The Role of Smallholder Farms in a Changing World In: y Paloma S, Riesgo L, Louhichi K, editors. The Role of Smallholder Farms in Food and Nutrition Security. Cham: Springer International Publishing. pp. 13–28. doi:10.1007/978-3-030-42148-9_2

FAOSTAT. 2021. FAOSTAT database collections. Food and Agriculture Organization of the United Nations. http://www.fao.org/faostat/

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol* **37**:4302–4315. doi:10.1002/joc.5086

Fitzpatrick MC, Keller SR. 2015. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecol Lett* **18**:1–16. doi:https://doi.org/10.1111/ele.12376

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. 2002. The structure of haplotype blocks in the human genome. *Science* **296**:2225–2229. doi:10.1126/science.1069424

Garcia-Erill G, Albrechtsen A. 2020. Evaluation of model fit of inferred admixture proportions. *Mol Ecol Resour* **20**:936–949. doi:10.1111/1755-0998.13171

Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R. 2014. ASReml User Guide Release 4.1 Functional Specification. *VSN Int Ltd, Hemel Hempstead, HP1 1ES, UK www.vsni.co.uk*.

Girma D, Assefa K, Chanyalew S, Cannarozzi G, Kuhlemeier C, Tadele Z. 2014. The origins and progress of genomics research on Tef (Eragrostis tef). *Plant Biotechnol J* **12**:534–540. doi:10.1111/pbi.12199

Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C. 2010. Food Security: The Challenge of Feeding 9 Billion People. *Science (80-)* **327**:812–818. doi:10.1126/SCIENCE.1185383

Griffith DA, Peres-Neto PR. 2006. Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology* **87**:2603–2613. doi:10.1890/0012-9658(2006)87[2603:smietf]2.0.co;2

Heffner EL, Sorrells ME, Jannink JL. 2009. Genomic selection for crop improvement. *Crop Sci*. doi:10.2135/cropsci2008.08.0512

Hijmans R, van Etten J. 2012. raster: Geographic analysis and modeling with raster data. *R Packag version 20-12*.

Hill WG, Weir BS. 1988. Variances and covariances of squared linkage disequilibria in finite populations. *Theor Popul Biol* **33**:54–78.

doi:10.1016/0040-5809(88)90004-4

Ho MD. 2011. International treaty on plant genetic resources for food and agriculturePlant Genetic Resources and Food Security. pp. 1–35. doi:10.4324/9781315530857-10

Ingram AL, Doyle JJ. 2003. The Origin and Evolution of Eragrostis tef (Poaceae) and Related Polyploids: Evidence from Nuclear Waxy and Plastid rps16. *Am J Bot* **90**:116–122.

IPCC PI sobre MC. 2017. Special Report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems (SR2), Ipcc.

Jamnadass R, Mumm RH, Hale I, Hendre P, Muchugi A, Dawson IK, Powell W, Graudal L, Yana-Shapiro H, Simons AJ, Van Deynze A. 2020. Enhancing {African} orphan crops with genomics. *Nat Genet* **52**:356–360. doi:10.1038/s41588-020-0601-x

Jia H, Li M, Li W, Liu L, Jian Y, Yang Z, Shen X, Ning Q, Du Y, Zhao R, Jackson D, Yang X, Zhang Z. 2020. A serine/threonine protein kinase encoding gene KERNEL NUMBER PER ROW6 regulates maize grain yield. *Nat Commun 2020 111* **11**:1–11. doi:10.1038/s41467-020-14746-7

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**:1403–1405. doi:10.1093/bioinformatics/btn129

Jones CD, Hughes JK, Bellouin N, Hardiman SC, Jones GS, Knight J, Liddicoat S, O'Connor FM, Andres RJ, Bell C, Boo KO, Bozzo A, Butchart N, Cadule P, Corbin KD, Doutriaux-Boucher M, Friedlingstein P, Gornall J, Gray L, Halloran PR, Hurtt G, Ingram WJ, Lamarque JF, Law RM, Meinshausen M, Osprey S, Palin EJ, Parsons Chini L, Raddatz T, Sanderson MG, Sellar AA, Schurer A, Valdes P, Wood N, Woodward S, Yoshioka M, Zerroukat M. 2011. The HadGEM2-ES implementation of CMIP5 centennial simulations. *Geosci Model Dev* **4**:543–570. doi:10.5194/gmd-4-543-2011

Juliana P, Poland J, Huerta-Espino J, Shrestha S, Crossa J, Crespo-Herrera L, Toledo FH, Govindan V, Mondal S, Kumar U, Bhavani S, Singh PK,

Randhawa MS, He X, Guzman C, Dreisigacker S, Rouse MN, Jin Y, Pérez-Rodríguez P, Montesinos-López OA, Singh D, Mokhlesur Rahman M, Marza F, Singh RP. 2019. Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. *Nat Genet* 1–10. doi:10.1038/s41588-019-0496-6

Ketema S. 1997. Eragrostis tef (Zucc.) Trotter: Promoting the conservation and use of underutilized and neglected crops. 12.

Khoury CK, Bjorkman AD, Dempewolf H, Ramirez-Villegas J, Guarino L, Jarvis A, Rieseberg LH, Struik PC. 2014. Increasing homogeneity in global food supplies and the implications for food security. *Proc Natl Acad Sci* **111**:4001–4006. doi:10.1073/PNAS.1313490111

Kidane YG, Mancini C, Mengistu DK, Frascaroli E, Fadda C, Pè ME, Dell'Acqua M. 2017. Genome wide association study to identify the genetic base of smallholder farmer preferences of durum wheat traits. *Front Plant Sci* **8**:1230. doi:10.3389/fpls.2017.01230

Kindt R, Coe R. 2005. Tree diversity analysis; A manual and software for common statistical methods for ecological and biodiversity studies, World. doi:10.13140/RG.2.1.1993.7684

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* **35**:1547–1549. doi:10.1093/molbev/msy096

Labeyrie V, Thomas M, Muthamia ZK, Leclerc C. 2016. Seed exchange networks, ethnicity, and sorghum diversity. *Proc Natl Acad Sci* **113**:98–103. doi:10.1073/pnas.1513238112

Lasky JR, Upadhyaya HD, Ramu P, Deshpande S, Hash CT, Bonnette J, Juenger TE, Hyma K, Acharya C, Mitchell SE, Buckler ES, Brenton Z, Kresovich S, Morris GP. 2015. Genome-environment associations in sorghum landraces predict adaptive traits. *Sci Adv* **1**:e1400218–e1400218. doi:10.1126/sciadv.1400218

Lemmon ZH, Reem NT, Dalrymple J, Soyk S, Swartwood KE, Rodriguez-Leal D,

Van Eck J, Lippman ZB. 2018. Rapid improvement of domestication traits in an orphan crop by genome editing. *Nat Plants* **4**:766–770. doi:10.1038/s41477-018-0259-x

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup 1000 Genome Project Data Processing, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079.

Liu X, Huang M, Fan B, Buckler ES, Zhang Z. 2016. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLOS Genet* **12**:e1005767. doi:10.1371/journal.pgen.1005767

Mancini C, Kidane YGYG, Mengistu DKDK, Pè MEMEME, Fadda C, Dell'Acqua M, Amit, Letegabriel, Tsedal, Tinasu B, Letekiros, Hadash, Letemedhin, Tinasu A, Kiros, Leteberhan, Letekidan, Hareya, Letecheal, Hiritu, Ametezigu, Amlak YGYG, Slassie PGGPGG, Girmay PGSPGS, Gebre GMM, Mesfin GSS, Desta K, Teklay S, Kidan HGHG, Egziabher TGTG, Desalegn PWPW, Gebre H, Micheal HGHG, Mebrahtu G, Hailemichael, Teklay A, Tesfay E, Gitehun A, Tadesse E, Asfaw M, Aragaw K, Brku T, Tadasse Y, Hailu M, Kassun A, Gel Aw G, Emagn M, Mitku F, Mengste A, Yigzaw B, Tesfaw E, Belay T, Yirga W, Sisay PAPA, Hailu B, Wale PTPT, Setegn M, Tesfie T, Meretie B, Kassie L, Biset T, Hailu Y, Agazu L, Yigzaw M, Yigzaw A, Abate G. 2017. Joining smallholder farmers' traditional knowledge with metric traits to select better varieties of Ethiopian wheat. *Sci Rep* **7**:13076. doi:10.1038/s41598-017-07628-4

Mascher M, Schreiber M, Scholz U, Graner A, Reif JC, Stein N. 2019. Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nat Genet* **51**:1076–1081. doi:10.1038/s41588-019-0443-6

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The

genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**:1297–1303. doi:10.1101/gr.107524.110

McSteen P, Malcomber S, Skirpan A, Lunde C, Wu X, Kellogg E, Hake S. 2007. barren inflorescence2 Encodes a Co-Ortholog of the PINOID Serine/Threonine Kinase and Is Required for Organogenesis during Inflorescence and Vegetative Development in Maize. *Plant Physiol* **144**:1000–1011. doi:10.1104/PP.107.098558

Mengistu DK, Mekonnen LS. 2012. Integrated Agronomic Crop Managements to Improve Tef Productivity Under Terminal Drought. *Water Stress*. doi:10.5772/30662

MoA. 2000. Agro-Ecological Zones of Ethiopia, MoA.

Occelli M, Mantino A, Ragaglini G, Dell'Acqua M, Fadda C, Pè ME, Nuvolari A. 2021. Traditional knowledge affects soil management ability of smallholder farmers in marginal areas. *Agron Sustain Dev* **41**:1–15. doi:10.1007/s13593-020-00664-x

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7**. doi:10.1371/journal.pone.0037135

Poland J. 2015. Breeding-assisted genomics. *Curr Opin Plant Biol* **24**:119–124. doi:10.1016/J.PBI.2015.02.009

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, De Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* **81**:559–575. doi:10.1086/519795

R Core Team. 2018. R: A language and environment for statistical computing.

Reuter HI, Nelson A, Jarvis A. 2007. An evaluation of void-filling interpolation methods for SRTM data. *Int J Geogr Inf Sci* **21**:983–1008. doi:10.1080/13658810601169899

Rhoné B, Defrance D, Berthouly-Salazar C, Mariac C, Cubry P, Couderc M, Dequincey A, Assoumanne A, Kane NA, Sultan B, Barnaud A, Vigouroux Y. 2020. Pearl millet genomic vulnerability to climate change in {West} {Africa} highlights the need for regional collaboration. *Nat Commun* **11**:5274. doi:10.1038/s41467-020-19066-4

Samberg LH, Shennan C, Zavaleta E. 2013. Farmer Seed Exchange and Crop Diversity in a Changing Agricultural Landscape in the Southern Highlands of Ethiopia. *Hum Ecol* **41**:477–485. doi:10.1007/s10745-013-9579-7

Schulzweida U. 2017. CDO User' s Guide. *Guide* 1–206. doi:10.5281/ZENODO.3539275

Shin J-H, Blay S, McNeney B, Graham J. 2006. LDheatmap: An R Function for Graphical Display of Pairwise Linkage Disequilibria Between Single Nucleotide Polymorphisms. *J Stat Softw* **16**:1–9. doi:10.18637/jss.v016.c03

Sloat LL, Davis SJ, Gerber JS, Moore FC, Ray DK, West PC, Mueller ND. 2020. Climate adaptation by crop migration. *Nat Commun 2020 111* **11**:1–9. doi:10.1038/s41467-020-15076-4

Stéphane Dray A, Bauman D, Blanchet G, Borcard D, Clappe S, Guenard G, Jombart T, Larocque G, Legendre P, Madi N, Wagner HH. 2021. Package "adespatial" Title Multivariate Multiscale Spatial Analysis. doi:10.1890/11-1183.1

Storey J, Bass A, Dabney A, Robinson D. 2021. Q-value estimation for false discovery rate control.

Tadele Z, Z T. 2019. Orphan crops: their importance and the urgency of improvement. *Planta* **250**:677–694. doi:10.1007/s00425-019-03210-6

Tajima F, Nei M. 1984. Estimation of evolutionary distance between nucleotide sequences. *Mol Biol Evol* **1**:269–285. doi:10.1093/oxfordjournals.molbev.a040317

Terlau W, Hirsch D, Blanke M. 2019. Smallholder farmers as a backbone for the implementation of the Sustainable Development Goals. *Sustain Dev* **27**:523–529. doi:10.1002/sd.1907

Turner HL, van Etten J, Firth D, Kosmidis I. 2020. Modelling rankings in R: the PlackettLuce package. *Comput Stat* **35**:1027–1057. doi:10.1007/s00180-020-00959-3

Valluru R, Reynolds MP, Davies WJ, Sukumaran S. 2017. Phenotypic and genome-wide association analysis of spike ethylene in diverse wheat genotypes under heat stress. *New Phytol* **214**:271–283. doi:10.1111/NPH.14367

Van Etten J, Beza E, Calderer L, Van Duijvendijk K, Fadda C, Fantahun B, Kidane YG, Van De Gevel J, Gupta A, Mengistu DK, Kiambi DAN, Mathur PN, Mercado L, Mittra S, Mollel MJ, Rosas JC, Steinke J, Suchini JG, Zimmerer KS. 2019. FIRST EXPERIENCES with A NOVEL FARMER CITIZEN SCIENCE APPROACH: CROWDSOURCING PARTICIPATORY VARIETY SELECTION THROUGH ON-FARM TRIADIC COMPARISONS of TECHNOLOGIES (TRICOT). *Exp Agric* **55**:275–296. doi:10.1017/S0014479716000739

van Etten J, de Sousa K, Aguilar A, Barrios M, Coto A, Dell'Acqua M, Fadda C, Gebrehawaryat Y, van de Gevel J, Gupta A, Kiros AY, Madriz B, Mathur P, Mengistu DK, Mercado L, Mohammed JN, Paliwal A, Pè ME, Quirós CF, Rosas JC, Sharma N, Singh SS, Solanki IS, Steinke J. 2019. Crop variety management for climate adaptation supported by citizen science. *Proc Natl Acad Sci U S A* **116**:4194–4199. doi:10.1073/pnas.1813720116

VanBuren R, Man Wai C, Wang X, Pardo J, Yocca AE, Wang H, Chaluvadi SR, Han G, Bryant D, Edger PP, Messing J, Sorrells ME, Mockler TC, Bennetzen JL, Michael TP. 2020. Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. *Nat Commun* **11**:1–11. doi:10.1038/s41467-020-14724-z

VanRaden PM. 2008. Efficient Methods to Compute Genomic Predictions. *J Dairy Sci* **91**:4414–4423. doi:10.3168/jds.2007-0980

Varshney RK, Bohra A, Yu J, Graner A, Zhang Q, Sorrells ME. 2021. Designing Future Crops: Genomics-Assisted Breeding Comes of Age. *Trends Plant Sci*

0. doi:10.1016/j.tplants.2021.03.010

Wang KL-C, Yoshida H, Lurin C, Ecker JR. 2004. Regulation of ethylene gas biosynthesis by the Arabidopsis ETO1 protein. *Nat 2004 4286986* **428**:945–950. doi:10.1038/nature02516

Weir BS, Cockerham CC. 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution (N Y)* **38**:1358. doi:10.2307/2408641

Weltzien E, Rattunde F, Christinck A, Isaacs K, Ashby J. 2019. Gender and Farmer Preferences for Varietal Traits. *Plant Breed Rev* 243–278. doi:10.1002/9781119616801.CH7

Woldeyohannes AB, Accotto C, Desta EA, Kidane YG, Fadda C, Pè ME, Dell'Acqua M. 2020. Current and projected eco-geographic adaptation and phenotypic diversity of Ethiopian teff (Eragrostis teff) across its cultivation range. *Agric Ecosyst Environ* **300**:107020. doi:10.1016/j.agee.2020.107020

Yerima ARIB, Achigan-Dako EG. 2021. A review of the orphan small grain cereals improvement with a comprehensive plan for genomics-assisted breeding of fonio millet in West Africa. *Plant Breed*. doi:10.1111/PBR.12930

Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan X, Zhu M, Zhao S, Li X, Liu X. 2020. {rMVP}: {A} {Memory}-efficient, {Visualization}-enhanced, and {Parallel}-accelerated tool for {Genome}-{Wide} {Association} {Study}. *bioRxiv* 2020.08.20.258491. doi:10.1101/2020.08.20.258491

Yıldız İ, Dy J, Erdoğmuş D, Kalpathy-Cramer J, Ostmo S, Campbell JP, Chiang MF, Ioannidis S. 2020. Fast and Accurate Ranking Regression 77–88.

Yu G, Smith DK, Zhu H, Guan Y, Lam TT. 2017. ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**:28–36. doi:10.1111/2041-210X.12628

Zhu Q, Smith SM, Ayele M, Yang L, Jogi A, Chaluvadi SR, Bennetzen J. 2012. High-throughput discovery of mutations in tef semi dwarfing genes by next-generation sequencing analysis. *Genetics* **192**:819–829.

Zhu Qihui, Smith SM, Ayele M, Yang L, Jogi A, Chaluvadi SR, Bennetzen JL.

2012. High-Throughput Discovery of Mutations in Tef Semi-Dwarfing Genes by Next-Generation Sequencing Analysis. *Genetics* **192**. doi:10.1534/GENETICS.112.144436

Tables

Table 1. Heritabilities (H^2) for farmers' participatory variety selection traits. H^2 values are given for each trait, type, and location combination. PA, panicle appreciation; OA, overall appreciation; M, men; W, women; ALL, measures combined by either type or location.

Trait	Type	Location	H^2
PA	ALL	ALL	0.48
PA	ALL	Adet	0.50
PA	ALL	Akaki	0.65
PA	M	ALL	0.43
PA	W	ALL	0.45
PA	M	Adet	0.35
PA	W	Adet	0.40
PA	M	Akaki	0.25
PA	W	Akaki	0.34
OA	ALL	ALL	0.81
OA	ALL	Adet	0.71
OA	ALL	Akaki	0.87
OA	M	ALL	0.74
OA	W	ALL	0.74
OA	M	Adet	0.53
OA	W	Adet	0.66
OA	M	Akaki	0.88
OA	W	Akaki	0.62

Table 2. Heritabilities (H^2) for agronomic traits. H^2 values are given for each trait, and location combination. DH, days to heading; DM, days to maturity; PH, plant height; PL, panicle length; PBPM, number of primary branches per main shoot panicle; TT, total tillers; CLF, first culm length; CDF, first culm diameter; PW, panicle weight; PY, panicle yield; GY, grain yield; BY, biomass yield; HI, harvest index; GFP, grain filling period; GFR, grain filling rate; BPR, biomass production rate; ALL, measures combined by location.

Trait	Location	H^2
DH	ALL	0.99
DH	Adet	0.96
DH	Akaki	0.97
DM	ALL	0.98
DM	Adet	0.90
DM	Akaki	0.89
PH	ALL	0.16
PH	Adet	0.92
PH	Akaki	0.90
PL	ALL	0.37
PL	Adet	0.88
PL	Akaki	0.81
PBPM	ALL	0.64
PBPM	Adet	0.83
PBPM	Akaki	0.83
TT	ALL	0.25
TT	Adet	0.65
TT	Akaki	0.60
CLF	ALL	0.13
CLF	Adet	0.88
CLF	Akaki	0.85
CDF	ALL	0.15
CDF	Adet	0.95
CDF	Akaki	0.96
PW	ALL	0.27
PW	Adet	0.90
PW	Akaki	0.88
PY	ALL	0.25
PY	Adet	0.92
PY	Akaki	0.91
GY	ALL	0.42
GY	Adet	0.92
GY	Akaki	0.93
BY	ALL	0.34
BY	Adet	0.84

BY	Akaki	0.80
HI	ALL	0.78
HI	Adet	0.68
HI	Akaki	0.76
GFP	ALL	0.96
GFP	Adet	0.79
GFP	Akaki	0.83
GFR	ALL	0.52
GFR	Adet	0.90
GFR	Akaki	0.92
BPR	ALL	0.32
BPR	Adet	0.80
BPR	Akaki	0.77

Table 3. Plackett-Luce estimates from farmer's overall appreciation (OA) of genotypes associated with genotypes' agronomic metrics DM, days to maturity; PH, plant height; PW, panicle weight; GY, grain yield; BY, biomass yield; GFP, grain filling period. The rankings were analyzed for the whole group (All) and in subsets among gender to assess differences in traits linkages within men and women farmers.

Group		Estimate	Std. Error	z value	Pr(>|z|)	
All	(Intercept)	-7.09	--	--	--	
	GY	0.000764	0.000135	5.66	1.51E-08	***
	DM	0.00625	0.00402	1.56	0.12	
	PH	-0.0123	0.00275	-4.46	8.17E-06	***
	PW	-0.0557	0.0716	-0.777	0.437	
	BY	-3.6E-05	2.72E-05	-1.32	0.186	
	GFP	0.011	0.00484	2.27	0.0233	*
Men	(Intercept)	-7.04	--	--	--	
	GY	0.000765	0.000175	4.36	1.29E-05	***
	DM	0.00542	0.00531	1.02	0.308	
	PH	-0.0114	0.00361	-3.17	0.00155	**
	PW	-0.0818	0.0932	-0.878	0.38	
	BY	-2E-05	0.000036	-0.562	0.574	
	GFP	0.00919	0.00637	1.44	0.149	
Women	(Intercept)	15.6	--	--	--	
	GY	-0.00084	0.00028	-3.01	0.00262	**
	DM	-0.186	0.0101	-18.5	< 2e-16	***
	PH	-0.139	0.00661	-20.9	< 2e-16	***
	PW	4.4	0.195	22.6	< 2e-16	***
	BY	0.000346	5.43E-05	6.38	1.74E-10	***
	GFP	-0.136	0.0104	-13.1	< 2e-16	***

Figures

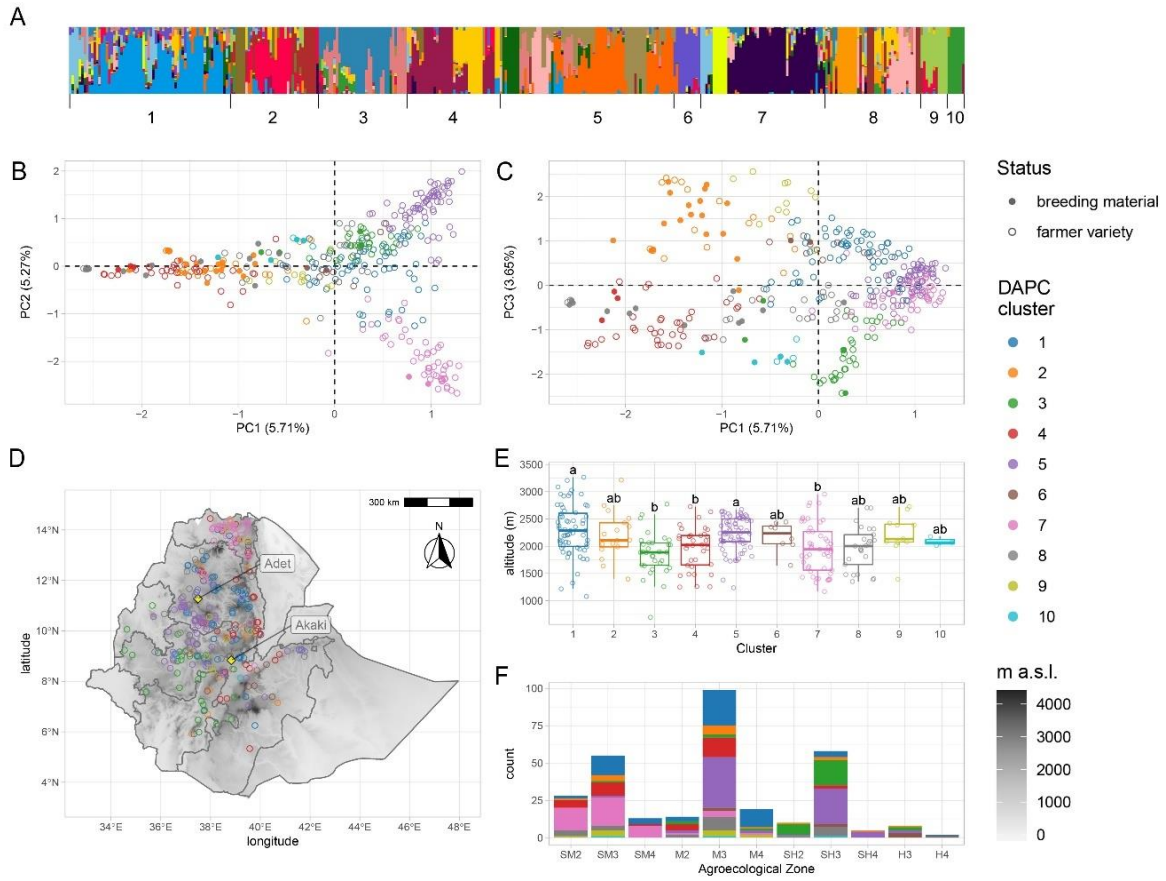

Figure 1. Genetic diversity of teff in Ethiopia. (**A**) ADMIXTURE results for the pruned SNPs dataset. Each vertical bar represents an individual, colored according to one of the 20 groups reported by the analysis. Bars are ordered according to the ten genetic clusters identified by DAPC, as reported by numbers on the x axis. (**B,C**) Principal Component Analysis of genome-wide SNPs. Taxa are colored according to their DAPC genetic cluster with colors according to legend. About 10.98% of the genetic diversity in the panel can be explained by the first two principal components, which clearly separate cluster 7 from clusters 2 and 4. Open and close circles represent farmer varieties and improved varieties, respectively. (**D**) Distribution of EtDP georeferenced landraces (N=314) across the altitudinal map of Ethiopia, color coded as in panel **B**. (**E**) Altitude distribution across the DAPC genetic clusters, shown by color, with letters on top of boxplots denoting significance levels with a pairwise Wilcoxon rank sum test with a Bonferroni correction for multiple testing. (**F**) Distribution of genetic clusters across agroecological zones of Ethiopia, with color coding as in panel **B.** SM2, warm sub-moist lowlands; SM3, tepid sub-moist mid-highlands; SM4, cool sub-moist mid-highlands; M2, warm moist lowlands; M3, tepid moist mid-highlands; M4, cool moist mid-highlands; SH2, warm sub-humid lowlands; SH3, tepid sup-humid mid-highlands; SH4, cool sub-humid mid-highlands; H3, tepid humid mid-highlands; H4, cool humid mid-highlands. This figure has six figure supplements.

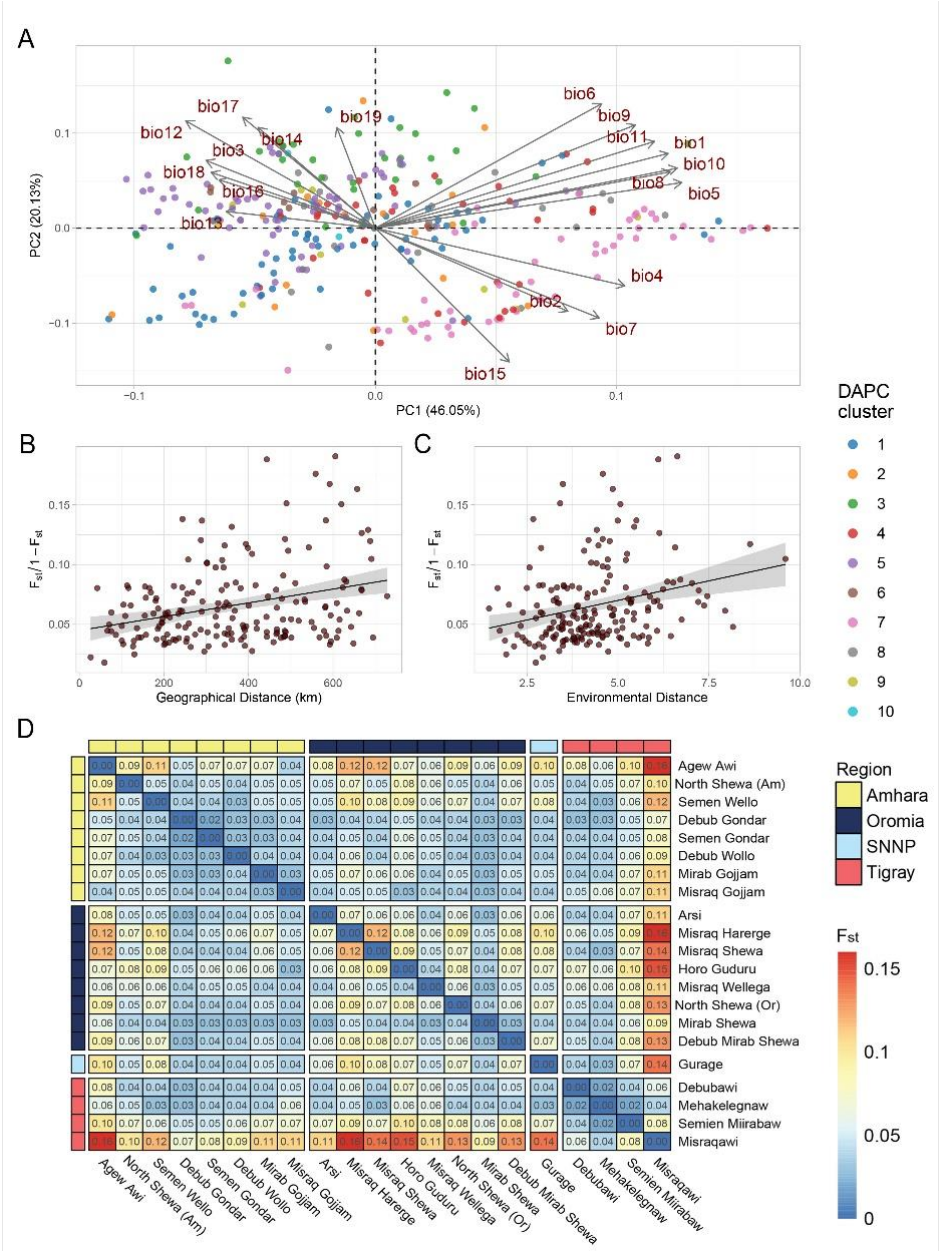

Figure 2. Teff diversity on the landscape. (**A**) Principal Component Analysis of bioclimatic diversity in the EtDP. Dots represent teff farmer varietiesf belonging to genetic clusters, colored according to legend. Vectors represent the scale, verse, and direction of bioclimatic drivers of teff differentiation. (**B**) Linear regression of F_{st} values in relation to geographic distance of accessions in the EtDP. Accessions were grouped by local district of sampling. (**C**) Linear regression of F_{st} values in relation to environmental distance of accessions, also grouped by district of sampling. (**E**) Pairwise F_{st} values between teff accessions grouped by local districts of sampling, as in (C) and (E). Local districts, *i.e.* sub-regional groups, are ordered by administrative regions according to legend. This figure has two figure supplements.

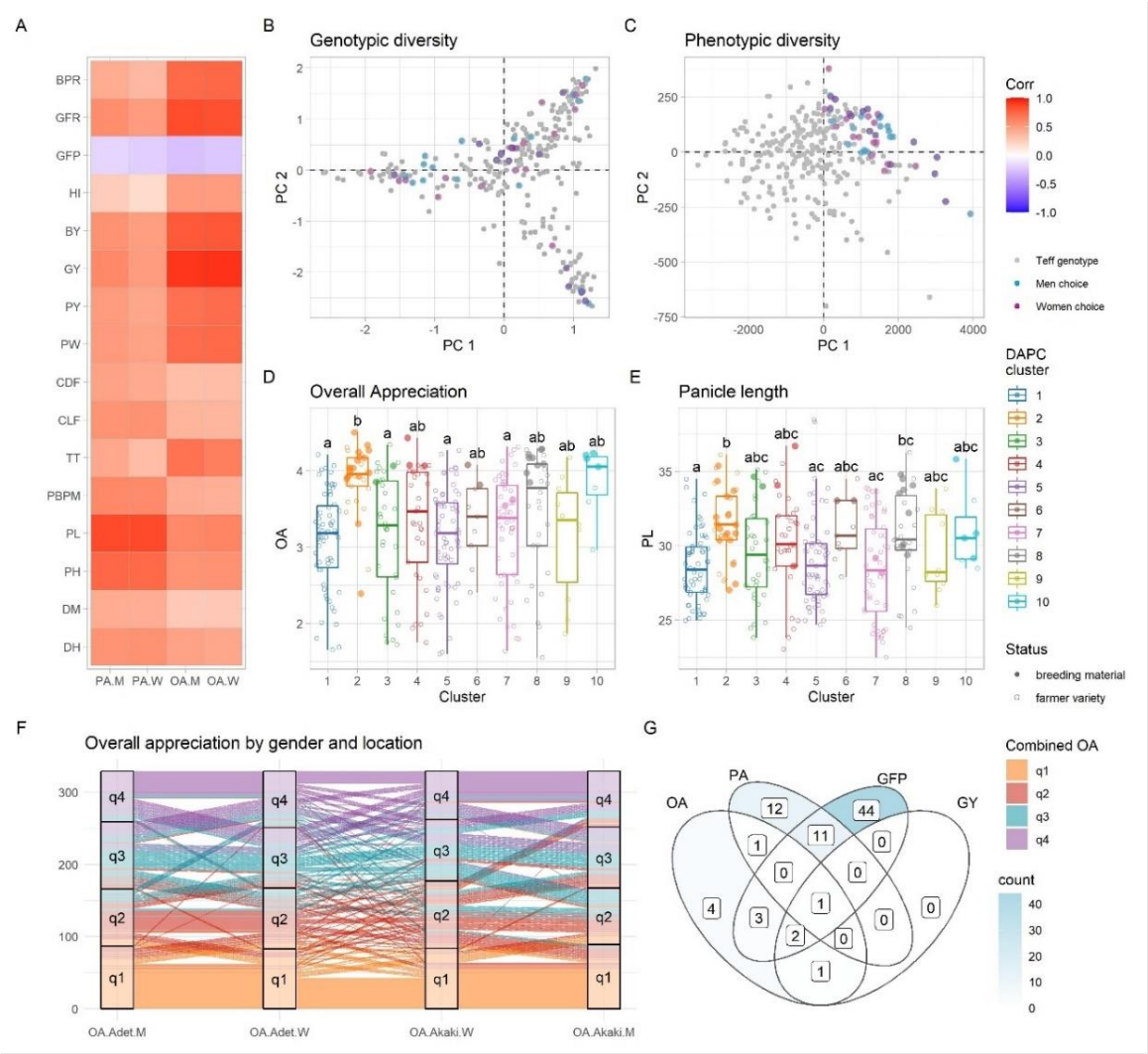

Fig. 3. Phenotypic diversity in the EtDP. (**A**) Pearson's correlations between agronomic traits (y axis) and farmer preference traits, by gender (x axis). Correlation values are expressed in color shades according to the legend to the right. (**B,C**) Top ranking genotypes (90th percentile of the OA distribution) selected by man (blue) and women (purple) farmers, overlaid to the EtDP diversity reported by a PCA summarizing SNP molecular markers (A) and phenotypes (C). Genotypes not selected are reported as gray dots. When men and women select the same genotype, the corresponding point appears in dark violet. (**D,E**) Trait distribution by DAPC cluster, for overall appreciation (D) and panicle length (E). Farmer varieties are represented by open circles, improved lines are represented by full circles. (**F**) Alluvial plot reporting the consistency of farmers' choice by quartiles of the OA distribution. Each vertical bar represents a combination of location (Adet, Akaki) and gender (M, W). EtDP accessions are ordered on the y axis according to their OA score in each combination. Alluvial flows are colored according to OA quartiles combined across gender and across location according to the legend (q1, q2, q3, q4). (**G**) Venn diagram reporting farmer varieties having values superior to the 75th percentile of the trait distribution of improved varieties (lower than the 25th percentile in the case of GFP). Each area of the Venn diagram reports the corresponding number of farmer varieties according to legend. The red star mark highlights a farmer varieties bearing a desirable combination of overall appreciation, panicle appreciation, grain filling period, and grain yield. DH, days to heading; DM, days to maturity; PH, plant height; PL,

panicle length; PBPM, number of primary branches per main shoot panicle; TT, total tillers; CLF, first culm length; CDF, first culm diameter; PW, panicle weight; PY, panicle yield; GY, grain yield; BY, biomass yield; HI, harvest index; GFP, grain filling period; GFR, grain filling rate; BPR, biomass production rate; OA, overall appreciation; PA, panicle appreciation. This figure has two figure supplements.

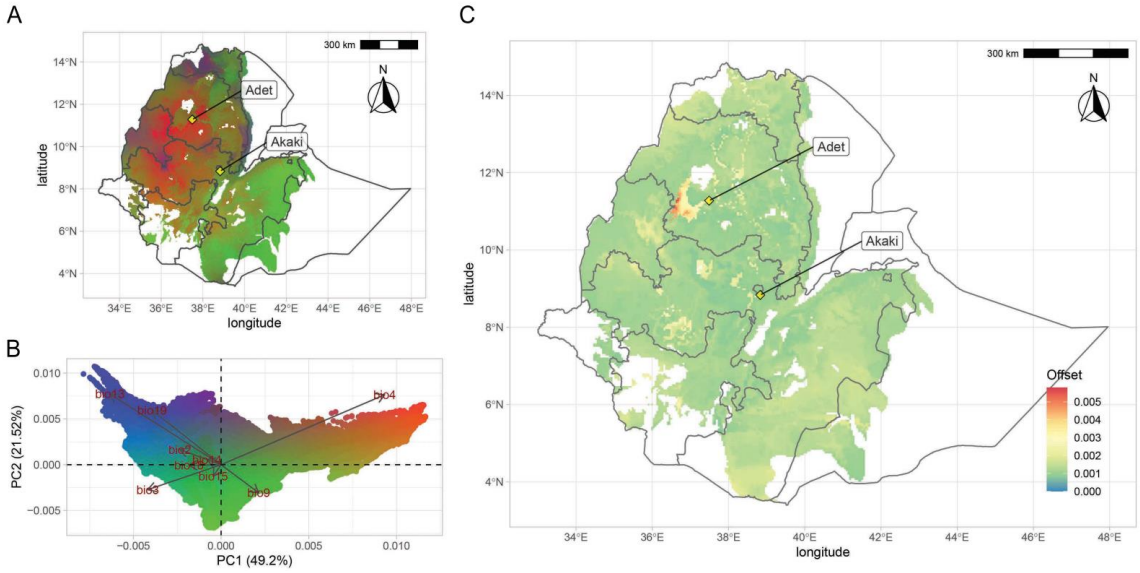

Figure 4. Teff genomic offset. (**A**) Geographic distribution of climate-driven allelic variation under current climates across the teff cropping area, with colors representing the three PC dimensions reported in panel (**B**). (**C**) Genomic vulnerability across the teff cropping area based on the RCP 8.5 climate projections. The color scale indicates the magnitude of the mismatch between current and projected climate-driven turnover in allele frequencies according to legend. Phenotyping locations are shown with yellow diamonds. This figure has seven figure supplements.

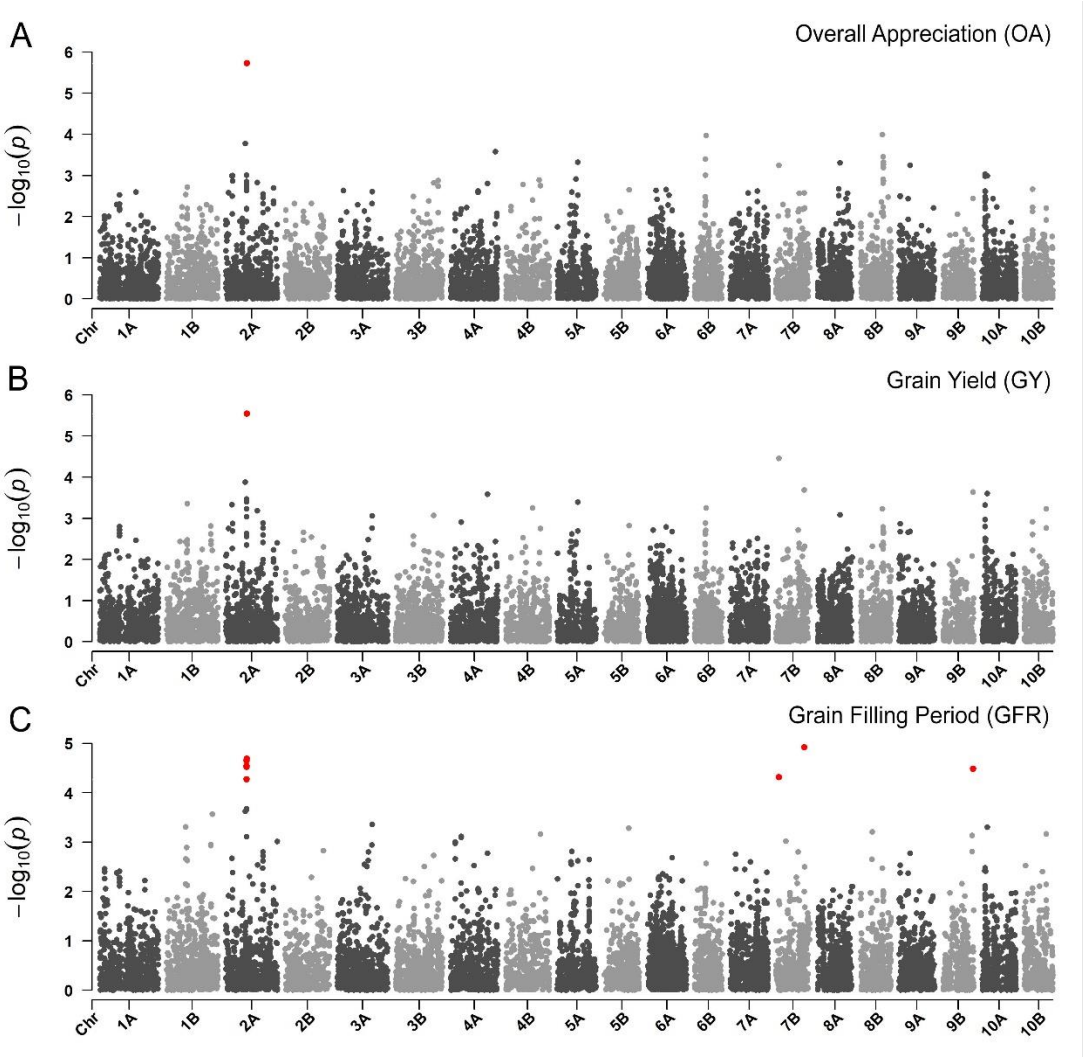

Figure 5 Manhattan plots reporting the GWAS result for (**A**) overall appreciation, (**B**) grain yield, and (**C**) grain filling period. On the x axis, the genomic position of markers. The y axis reports the strength of the association signal. SNPs are ordered by physical position and grouped by chromosome. Quantitative trait nucleotides (QTNs), *e.g.* SNPs surpassing a threshold based on a False Discovery Rate of 0.05, are highlighted in red. A strong signal on chromosome 2A matches across PVS and metric traits. This figure has two supplements.