Data-driven, participatory characterization of traditional farmer varieties discloses teff (*Eragrostis tef*) adaptive and breeding potential under current and future climates

Aemiro Bezabih Woldeyohannes[& 1,2], Sessen Daniel Iohannes[& 1], Mara Miculan[1], Leonardo Caproni[1], Jemal Seid Ahmed[1], Ermias Abate Desta[2], Carlo Fadda[3], Mario Enrico Pè[1], Matteo Dell'Acqua[1]*

[1] Institute of Life Sciences, Scuola Superiore Sant'Anna, Pisa, Italy

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

[3] Alliance of Bioversity and CIAT, Rome, Italy

[&]Equal contributing authors

* Correspondence to: Matteo Dell'Acqua, m.dellacqua@santannapisa.it

Abstract

In smallholder farming systems, traditional farmer varieties of neglected and underutilized crops species (NUS) support the livelihoods of millions of growers and consumers. NUS combine cultural and agronomic value with local adaptation, and call for transdisciplinary methods to evaluate their breeding potential. Here, we combined farmers' traditional knowledge, genomics, and climate science to characterize 366 Ethiopian teff (*Eragrostis tef*) farmer varieties and breeding materials. We found that teff genetic diversity in Ethiopia could be organized in six genetic clusters associated to climate variation on the landscape. A participatory evaluation conducted in collaboration with local farmers could consistently identify best performing varieties and inform a genome wide association study to identify candidate genes for farmers' appreciation, phenology, yield, and local adaptation. By modelling the genomic adaptation of teff to current and projected climates, we identified an area around lake Tana where teff cropping will 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

Main

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[1], and the overall composition of food systems is uniform worldwide[2]. Yet, hundreds of neglected or underutilized crop species (NUS) are actively cultivated in highly diversified, small-scale cropping systems, where they support the livelihoods of millions of people[3,4]. NUS are species that experienced scant research and breeding improvement. Not only NUS diversity is a proxy of pedoclimatic diversification of cropping systems, but it is also a telling of socioeconomic diversity and cultural heritage of local farmers[5]. Rich NUS agrobiodiversity is conserved *in situ* in smallholder agriculture systems, where the selection and cultivation conducted by local farmers resulted in the development 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 thus unlock the potential of NUS towards the sustainable intensification of farming systems, both in challenging cropping environments and in large-scale agricultural systems and associated markets[3,6,7].

The research community can now leverage the big data revolution to bridge the gap between NUS and 21st century agriculture[6,8]. Genomic tools allow us to rapidly characterize large germplasm collections and to identify genetic factors responsible for traits of agronomic interest[9,10], unlocking desirable agrobiodiversity to breeding[11] and enabling genomic selection to accelerate genetic gains[12]. Genomic data can be put in relation with increasingly precise current and projected climatic data and derive information not only on locus-specific adaptation[13], but also on genomic vulnerability under climate change scenarios[14–16]. Data-driven methods can also be applied to characterize the socioeconomic contexts in which crops are grown[17], generating information that is critical to understand cropping dynamics in smallholder farming systems[18,19]. Participatory varietal selection approaches, which harness farmers' experience on local agricultural diversity and agronomic potential, directly involve farmers in the evaluation of breeding materials[20,21], and can be combined with genomic data to identify genomic loci responsible for farmers' appreciation[22] and model local crop performance in farmer fields[23].

The Ethiopian farming system is a paradigm of challenging agricultural ecosystems where NUS farmer varieties are widely cultivated, and where an untapped potential for sustainable food systems and agricultural intensification of local agriculture. In Ethiopia, 85 million people live in rural areas, most of which are subsistence-based smallholder farmers that are responsible for about 90% of the cultivated land and agricultural output[24]. Teff (*Eragrostis tef*), a tetraploid NUS belonging to the *Chloridoideae*

subfamily, is a staple crop in the whole horn of Africa, where it has been cultivated for millennia and used for traditional preparations including *enjera* and *tella,* and is increasingly prized as super food in western markets[25,26]. Research in teff is rapidly evolving; a draft genome sequence[27]and a high quality genome sequence[28]recently brought it to the international genomics research spotlight. Breeding efforts have been underway since decades, and segregant and mutagenized populations are also available[29,30]. Yet, the full potential of teff is still undisclosed, and teff yields remain still much lower than potentially attainable and substantially lower than those of other cereals grown in the region[31–33]. However, thousands of locally adapted teff landraces show enormous potential in environmental adaptation and phenotypic diversity[34]. Combining data-driven research approaches in a transdisciplinary characterization may accelerate teff breeding efforts to unlock its full potential.

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 in two locations in Ethiopia. Fifteen women and twenty men experienced teff farmers were asked to evaluate the teff genotypes, providing quantitative information that could allow to prioritize genetic materials that are best adapted to local environments and that are able to meet local agricultural needs. 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 all sources of information in a genome-wide association study framework to identify genomic loci with relevance for adaptation, performance, and farmers' preferences, and we discuss candidate genes for teff improvement. We conclude discussing the potential of data-driven participatory approaches to characterize NUS diversity valorizing their heritage and 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 farmer varieties purified from landraces 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 (Supplementary Figure 1, Supplementary Table 1). The genomic diversity of the EtDP was assessed using 12,153 high-quality, genome wide single nucleotide polymorphisms (SNPs) derived from the DNA sequencing 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 (Supplementary Table 2). Chromosomes showed consistently higher pericentromeric LD, with localized LD peaks in telomeric regions (Supplementary Figure 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[28].

The internal administrative borders of Ethiopia are markers of cultural and historical diversity, and this is reflected in the genetic diversity of teff landraces cultivated by local farmers. Teff accessions from Tigray, in the north of Ethiopia, were markedly separated from the rest (Supplementary Figure 3a). Although teff breeding lines could be put in relation with landraces sampled from across the country, they had a relatively narrow genetic base that failed to sample the broad diversity available in the EtDP (Supplementary Figure 3b). EtDP accessions showed varying degrees of genetic admixture (Fig. 1a) and could be grouped in 6 non-overlapping clusters (Supplementary Figure 4). About 15% of the genetic variability in the teff panel could be explained by the first three principal components (PCs) of SNP data, which neatly separated cluster 1 from cluster 2 and cluster 4 (Fig. 1b-c).

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[35], or establishing breeding groups to explore heterosis potential in teff[36]. The preservation and valorization of gene pools supports the capacity to respond to future needs and priorities in the face climate change and the dynamic consumers preferences. Currently, teff improved lines predominantly come from genetic clusters 2, 4, and 5. A group of accessions sampled in North-Eastern Tigray was very different from the rest of the collection (Fig. 1d), and mostly belonged to genetic cluster 1. EtDP genetic clusters may also support teff germplasm conservation efforts. We found that altitude and agroecological zones could tell apart genetic clusters, with different ancestries succeeding one another from hot and dry areas to cold and humid areas (Fig. 1e-f). Teff is mostly grown in areas characterized by relatively low humidity[34]. Although it shows lower water requirements than other cereals such as barley, teff has a moderately sensitive response to water stress and is likely to provide higher yields when water supply is optimal[37].

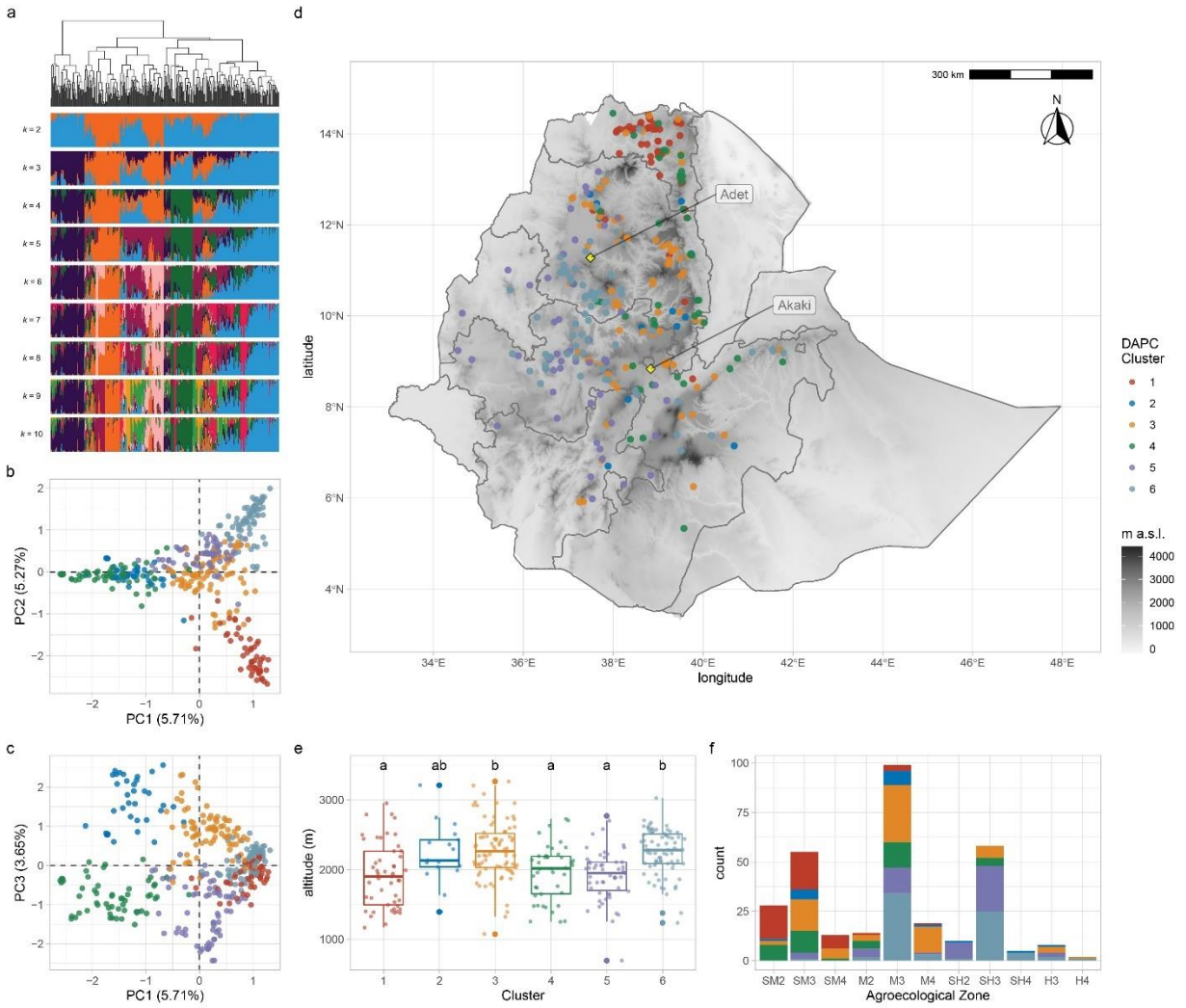

Figure 1. Genetic diversity of teff in Ethiopia. (**a**) ADMIXTURE results for the pruned SNPs dataset at values of K ranging from 2 to 10 (**b,c**) Principal Component Analysis of genome-wide SNPs. Taxa are colored according to their relative genetic cluster computed using the Discriminant Analysis of Principal Components (DAPC). About 10.98% of the genetic diversity in the panel can be explained by the first two principal components, which clearly separate cluster 1 from clusters 2 and 4. (**d**) Distribution of EtDP georeferenced landraces (N=314) across the altitudinal map of Ethiopia, color coded as in panel **b**. (**e**) Altitudinal distribution of genetic clusters, shown by color, with letters on top of boxplots denoting significance levels with a pairwise Wilcoxon rank sum test with Bonferroni correction for multiple testing. (**f**) Genetic cluster composition of 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.

The distribution of teff genetic variation is associated with geographic and environmental factors

Landraces evolve at the interface of natural and anthropogenic selection[38], hence we hypothesized that teff genetic clusters might be associated with local environmental conditions. High levels of admixture indicated limited population stratification (Supplementary Figure 5), yet significant associations could be observed between genetic clusters and agroecological zones of Ethiopia (Chi-square test, p-value < 2.2e-16; Supplementary Figure 6). Climatic indicators could be summarized by PCs computed from bioclimatic variables, identifying precipitation and temperature gradients as well as seasonal patterns (Fig. 2a-b). Genetic cluster 1 was mainly distributed in warmer and drier climates, while cluster 5 and 6 came from colder and wetter areas (Supplementary Figure 7). Extant landraces diversity is not only contributed by climate, but also by seed circulation. Smallholder farmers are connected in formal and informal seed exchange networks[39] which are drivers of genetic diversity[40,41]. The district of sampling of teff landraces in the EtDP, a proxy of geography and ethnicity in Ethiopia, was used to aggregate accessions and calculate genetic distances as a measure of fixation index (F_{st}). F_{st} values were significantly associated with geographic distance (Mantel r = 0.31; p = 9e-04) and environmental distance (Mantel r =0.352; p =0.0137), showing that a combination of isolation by distance and environmental adaptation might have shaped teff genetic diversity (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[42] and archaeobotanical excavations reported its cultivation since the Pre Aksumite period (before I Century CE)[26]. 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 grouped near to genetic cluster 3 (Supplementary Figure 8a), lacking a precise geographic provenance (Supplementary Figure 8b) but being more frequent in cool moist mid-highlands (Supplementary Figure 6).

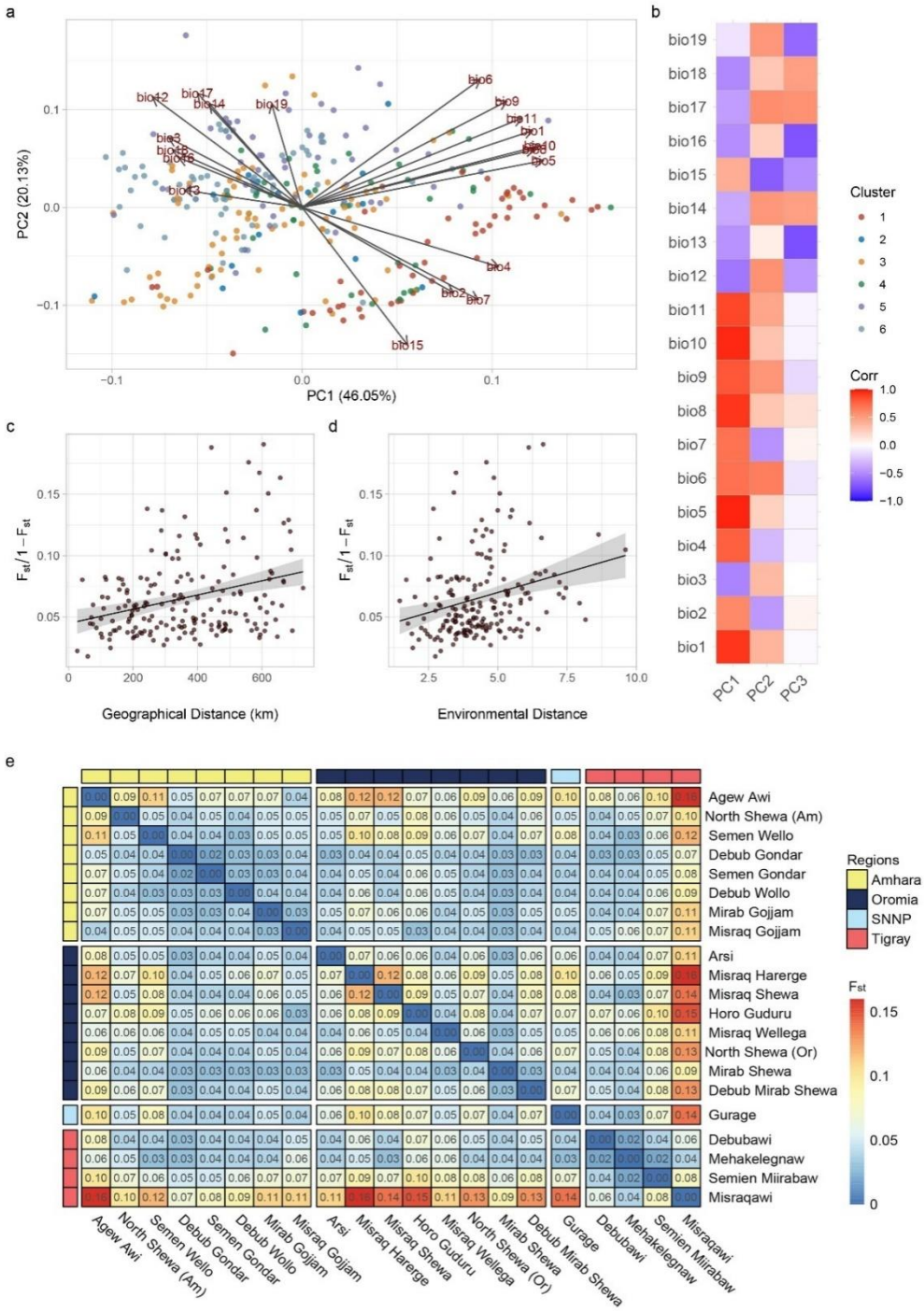

Figure 2. Teff diversity on the landscape.(**a**) Principal Component Analysis of bioclimatic diversity in the EtDP. Dots represent teff landraces colored according to genetic clusters, as in legend. Vectors represent the scale, verse, and direction of bioclimatic drivers of teff differentiation. (**b**) Correlation between derived bioclimatic PCs and original bioclimatic variables, colored according to legend. (**c**) Evolution of F_{st} values in relation to geographic distance between teff accessions grouped by sub-regional administrative borders of Ethiopia. (**d**)Evolution of F_{st} values in relation to environmental distance of teff accessions grouped by sub-regional administrative borders of Ethiopia. (**e**) Pairwise F_{st} values between teff accessions grouped by sub-regional administrative borders of Ethiopia. Sub-regional groups of samples are ordered by administrative regions according to legend.

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 high potential for teff cultivation. Men and women farmers expert teff growers evaluated individual plots, while researchers collected metric traits. Farmer evaluations were highly repeatable and had a strong genetic basis. The overall appreciation provided by farmers across genders and across locations had a broad sense heritability (H^2) of 0.81 (Supplementary Table 3). H^2 of grain yield combined across the same two locations was of 0.42 (Supplementary Table 4). Farmers' appreciation of teff genotypes was positively correlated with yield and yield components, most notably panicle traits, biomass, and grain filling rate (Fig. 3a, Supplementary Figure 9). Previous studies showed that farmers' appreciation is genetically determined, and may be used to perform genomic prediction[23] and identify genomic associations[22]. The high H^2 achieved by farmers' overall appreciation may be due to the fact that farmers, in providing their overall evaluation, not only consider yield but also yield component traits with high heritability.

The top ranking teff accessions according to men and women farmers captured different genetic backgrounds (Fig. 3b), but the same trait combinations (Fig. 3c), indicating that farmers were consistently preferring the same teff types regardless of genetic background and geographic provenance, *i.e.* high yielding, high-biomass and fast maturing landraces. Teff improved varieties showed high performance for farmers' overall appreciation (OA) as well as for agronomic traits (Fig. 3d-e). This is not a surprise since the evaluation was conducted in high potential areas, which are optimal environment for most teff breeding materials. Men and women farmers could efficiently identify most desirable varieties with high consistency, regardless of gender and location (Fig.3f). This is in line with previous literature, that shown that smallholder farmers' preference of men and women alike are quantitative and repeatable[22,43]. Several landraces from different genetic backgrounds recorded similar, at times superior, performances than improved lines (Fig. 3g). Accessions belonging to genetic clusters 2, 4 and 5 display longer days to heading and days to maturity, higher plant heights and panicle lengths, greater number of total tillers and higher yields (Supplementary Figure 10). By selecting landraces that outperform improved varieties' performances in target traits for breeding, it is possible to prioritize landrace accessions for teff improvement (Fig. 3e) or even immediately make these landraces available to farmers, as suggested by previous experiences in wheat[44]. Short maturation time is paramount to achieve harvestable yield in areas exposed to terminal drought[45], 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 elevated potential yield. We found that many landraces had a shorter grain filling period than most improved lines, and that landrace EBI *9551* combined this trait with superior yield and farmers' appreciation (Fig. 3g).

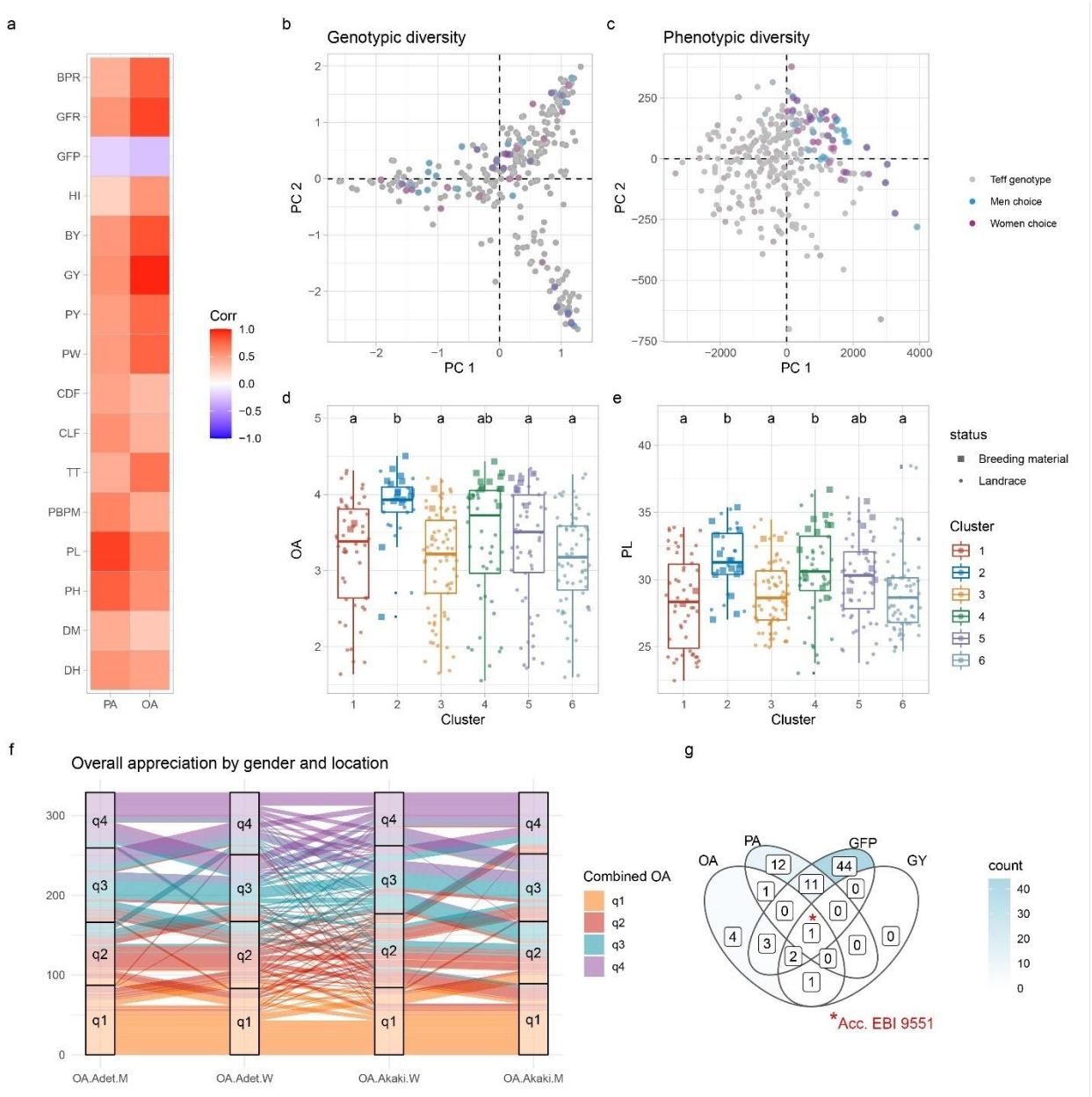

Fig. 3. Phenotypic diversity in the EtDP. (**a**) Pearson's correlations between agronomic traits (y axis) and farmer preference traits (x axis) in the teff collection. Correlation values are expressed in color shades according to the legend to the right. (**b,c**) Top ranking genotypes (90th percentile of the OA distribution) selected by man (blue) and women (purple) farmers, overlaid to the EtDP diversity reported by a Principal Components Analysis summarizing either SNP molecular markers (b) or phenotypes (c). Genotypes not selected are reported as gray

dots. Due to partial transparency of coloring, when men and women select the same genotype, the corresponding point appears in dark violet. (**d,e**) Trait distribution by genetic cluster, for overall appreciation (d) and panicle length (e). Landraces are represented by dots; improved lines are represented by squares with colors according to legend. (**f**) Alluvial plot reporting the consistency of farmers' choice by quartiles of the overall appreciation 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 landraces 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 landraces according to legend. The red star mark highlights a landrace 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.

Participatory, climatic, and agronomic diversity identify candidate loci for teff breeding

The data deriving from the transdisciplinary characterization of the EtDP can be integrated to highlight interactions and emerging properties in the collection, identifying genomic loci associated with agronomic performance, local adaptation, and farmer preferences. This can be achieved in a genome wide association study (GWAS) framework to establish a lasting toolbox to support teff improvement via either marker assisted selection, genomic selection, or new breeding technologies. A GWAS led to the identification of a total of 193 unique quantitative trait nucleotides (QTNs) (Supplementary Table 5). A hundred and one unique QTNs were associated with bioclimatic indicators at sampling sites. Seventeen more QTN were associated with farmers' appreciation. 115 QTNs could be grouped in 59 haplotype blocks marking the co-inheritance of contiguous SNPs. This information could be used to guide marker assisted breeding via allele pyramiding, but the recent teff genome annotation[28] allows to identify suggestive candidate genes underlying teff phenotypes. We focused on these LD blocks to identify sequence homology of predicted teff proteins with protein sequences of *Arabidopsis thaliana* and *Zea mays* (Supplementary Table 6).

We identified two loci on chromosome 6B associated with overall appreciation and panicle length (lcl|6B-7907481, lcl|6B-7907767). The corresponding LD block includes 24 gene models, some of which have predicted products with sequence homology with proteins with known phenotypes. *Et_6B_049642* encodes for a putative phosphoinositide phosphatase with strong homology with the product of *A. thaliana AT3G51460*. In Arabidopsis, knockout mutants of this gene show defective root hair development[46]. The product of *Et_6B_050159*, in the same LD block, shows sequence homology with

that of the *Zea mays* O-methyltransferase *ZRP4* gene, which has been reported to accumulate in roots and contribute to suberin biosynthesis[47,48]. We identified a locus on chromosome 2A that was associated with grain yield, grain filling rate and overall appreciation (lcl|2A-14415768). 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. *Et_2A_015515*, also in the block, is a homolog of a maize Serine/threonine protein kinase 3, belonging to abroad class of proteins that was associated with inflorescence development[49] and grain yield[50] in maize.

Teff cultivation is vulnerable to climate change

The extant genomic diversity of teff may be ill adapted to future climate scenarios. We used a gradient forest (GF) machine learning algorithm to calculate the importance of environmental gradients on genomic variation across the landscape and to estimate the genomic offset of teff under projected climates. Teff allele frequencies turnover across the cropping area was best predicted by the geographic distribution of accessions (through Moran's eigenvector map variables, MEM) and by precipitation indicators, particularly precipitation of the coldest quarter (bio19) and precipitation of the wettest month (bio13) (Supplementary Figure 11). This is in agreement with observed F_{st} patterns (Fig. 2c-d). The GF allowed to model the interaction of historical climate data with current teff allelic frequencies, identifying climate-driven genomic variation patterns across the landscape (Fig. 4a-b). Approximately a quarter of the SNPs (3,049 in 521 LD blocks) had predictive power towards the GF: of these, 176 showed F_{st} values in the 99th percentile of the distribution (Supplementary Table 7).

The GF could be linked to GWAS to identify genomic loci associated to climate, agronomic performance, farmers' preference, and adaptive potential of teff (Supplementary Figure 12). Not surprisingly, LD blocks containing QTNs for days to maturity are associated with the GF model, supporting the importance of phenology in teff adaptation and geographic distribution[34]. 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. This LD block harbors 176 gene models among which four produce proteins with high homology with maize and Arabidopsis proteins (Supplementary Table 6). The predicted product of *Et_1A_007229* shows a remarkable similarity with a phosphatidylinositol kinase that is involved in flower development and has been shown to influence floral transition in condition of abiotic stress[51]. Days to maturity was also associated with a LD block at

20.8 Mb on chromosome 6A that was predicted by the GF model. In this block, nine predicted gene products share homology with Arabidopsis and maize (Supplementary Table 6). The protein encoded by *Et_6A_046800* has high homology with that produced by*AT4G02680*, an ETO1-like protein involved in the regulation of ethylene synthesis[52]. Ethylene is a key plant hormone that has been shown to be related to spike development and senescence[53,54].

These candidate genes are not yet validated. With increasing availability of teff information, corroborated by reverse genetic approaches enabled by mutagenized populations[55], it will be possible to validate candidate genes underlying traits of interest. Teff breeding could then fully benefit of targeted editing[56] to speed up the development of new varieties with improved yield, local adaptation and adherence to local preferences.

The teff adaptive potential across the landscape varied in magnitude and distribution according to different predicted climate scenarios for 2070 (Supplementary Figure 13). We then computed the genomic-adaptive offset between current and future climate scenarios to identify vulnerable areas (Fig. 4c, Supplementary Figure 14). In all representative concentration pathways, the highest offset was predicted in the north-western highlands of the Amhara region, south of lake Tana. Compared to other regions of the country, we found a decreasing trend of rainfall change in this region across all emission scenarios (Supplementary Figure 15). 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 (Supplementary Figure 16). Decreasing trends of rainfall during the main growing season are predicted in all projected scenarios, suggesting the possibility of a season shifting that might be especially critical for teff development stages (Supplementary Figure 17).

A valid adaptation strategy could be the assisted migration of teff genotypes from areas of higher vulnerability to areas of lower vulnerability[14], although crop migration and varietal replacement strategies need to take into account ecological and socio-economic factors, including the impacts on existing ecosystems and on farmers' adoption of migrated varieties[57]. Collating genetic, climatic, and participatory data, teff breeding efforts may effectively anticipate what lies ahead.

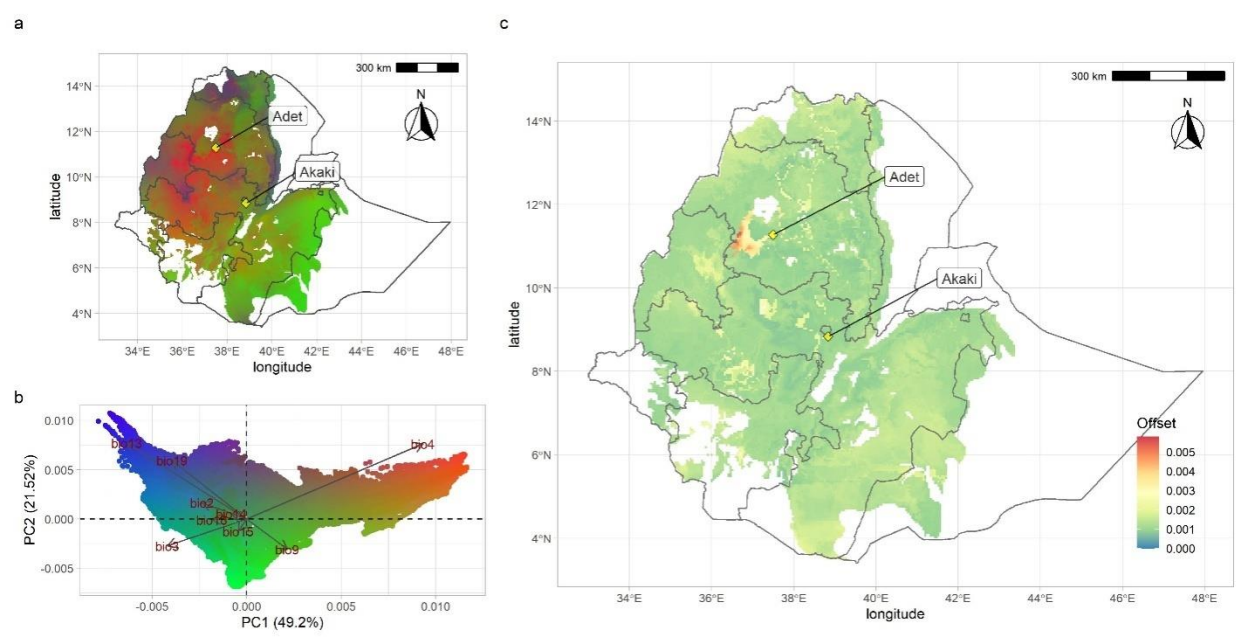

Figure 4. Teff genomic offset. (**a-b**) Measures of genetic differentiation of teff by either geographic distance (a) or climatic distance (b) between groups of samples. (**c-d**) Contribution to the GF model accuracy by measures of spatial structure (c, MEM) and by measures of climatic variation (d). Blue bars represent variables related to rainfall, red bars represent variables related to temperature. (**e**) 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 (**f**). (**g**) 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

Conclusion

A comprehensive interpretation of crop performance is the 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[58]. The success of crop varieties is not only determined by yield performance, but also by adaptation to local agricultural needs and growing conditions[59,60]. The integration of genomic, climatic, and phenotyping diversity in a participatory framework may help tailoring 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[23,61]. Here, farmers' knowledge is integrated in a broader picture considering climatic adaptation. Transdisciplinary methods may support the integration of smallholder farmers in modern breeding and agricultural value chains. Modern data-driven research

may efficiently harness the diversity produced by the incessant selection with which farmers shaped current agrobiodiversity. The combination of this information with that produced by genomics and climate reanalysis provides breeding with additional information that can be projected to broader temporal and spatial scales[62], enabling predictive varietal improvement.

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[63]. Temperature increases are also expected to result in more intense heat waves and higher evapotranspiration rates, which coupled with the projected alteration of rainfall patterns will affect multiple aspects of local economic development and 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[64], but our analysis show that their characterization is at hand. Teff is rapidly emerging from the NUS status and is projected towards modern breeding approaches. The same fate is being followed by other NUS including fonio[65], proso millet[66], *Amaranthus*[67]. Genebank genomics can unlock the rapid characterization of *ex situ* NUS agrobiodiveristy[10], which in teff could be indicative of even more diversity yet to be sampled in farmer fields[34]. Decentralized varietal evaluation approaches may then greatly accelerate the systematic testing of these genetic resources[21,23] to produce new varieties with higher local adaptation resulting in higher farmers' varietal adoption[68].

The characterization of larger diversity panels, coupled with improved genomic tools, is needed to fully support next generation breeding technologies[69] and large-scale genomic selection[11] to produce new and improved modern varieties. The teff genome annotation is yet to be fully refined and integrated in comparative genomics databases, but increasing interest in the crop will soon further enrich the teff breeding toolbox. 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 and DNA extraction

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 et al[34].

Landraces from the EBI were purified selecting and reproducing one single panicle representative of each accession. For the scope of this paper, we define *farmer varieties* the uniform genotypes derived from the purification of *ex-situ* accessions (*i.e.* landraces). *Farmer varieties* are a proxy of landraces originally collected in farmer fields, and are discussed as such. The EtDP includes 321 landraces which are a representative sample of the larger EBI collection in terms of geographic distribution and phenotypic diversity. EtDP landraces are 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 (Supplementary Table 1). 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.

Sequencing and variant calling

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)[70]. 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[71] 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)[28]with BWA (Burrows-Weeler-Aligner v.0.7.12) using the MEM algorithm with standard parameters[72]. Alignments were sorted and indexed with PicardTools (http://broadinstitute.github.io/picard/) and samtools[73].

Single nucleotide variants were identified with GATK[74]HaplotypeCaller algorithm (version 4.2.0), run in per-sample mode followed by a joint genotyping step completed by GenotypeVCFstool. 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%. Accessions having more than 20% of missing data were discarded. Statistics were calculated with *vcftools*[75].

Spatial and bioclimatic characterization

GPS coordinates of EtDP teff landraces were derived from EBI passport data and projected onto the map of Ethiopia using the R/raster[76]. Altitudes were assigned to each landrace based on GPS coordinates, using the CGIAR SRTM database at 90 m resolution[77]. GPS coordinates were projected onto the agroecological-zones map of Ethiopia provided by the Ethiopian Institute of Agricultural Research (EIAR)[78], which subdivides Ethiopia into different zones according to altitudinal ranges and temperature and rainfall patterns. Teff accessions were mapped into the regional administrative boundaries of Ethiopia using the data downloaded from the database of global administrative boundaries[79]. Current climate data (1970-2000 averages) and climate projections relative to teff landraces' sampling sites were retrieved from the WorldClim 2 database of global interpolated climate data[80] at the highest available spatial resolution, using R/raster. Nineteen bioclimatic indicators representing annual trends, seasonality and extremes or limiting environmental factors were considered. Collinearity among historical bioclimatic variables was previously checked with the *ensemble.VIF()* function in R/BiodiversityR[81]. 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)[82] 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.

Participatory evaluation and phenotyping

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). Accessions were planted in two replications per site using alpha lattice designs, in plots consisting of three rows of 1m in length and 0.2m 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. The following 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[34].

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. Farmers were engaged in focus group discussions prior to the PVS to discuss most relevant traits in teff and to attend training on the PVS. During the evaluation, farmers were divided in gender-homogeneous groups with five people each. Groups were conducted across the field from random entry point and asked to evaluate two teff 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 [trait] 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. PVS was conducted close to physiological maturity in each location so to maximize variation between plots.

Genetic diversity analyses

Phylogenetic relationships in the EtDP were assessed on a pruned set of SNP markers with MAF > 0.05. Pruning was performed with the PLINK[83]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[84]. A neighbor joining (NJ) tree was developed computing genetic distances using the Tajima-Nei method[85], performing 500 bootstrap resampling, using MEGA X [86]. Different NJ tree visualizations were produced using R/ggtree[87]. A principal component analysis (PCA) and a discriminant analysis of principal components (DAPC) were performed with R/adegenet[88]. 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[89] 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[90].

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[91]. The LD was plotted against physical positions, averaging pairwise r^2values 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 20 chromosomes according to the Hill and Weir[92] 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[93].

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 previously described in *sequencing and variant calling* section.

Phenotypic diversity analyses

Best linear unbiased predictions (BLUP) of agronomic and PVS traits were computed with R/ASReml[94]. BLUPs for agronomic traits were derived from the general model in Eq. (1):

$$y_{ijkn} = \mu + g_i + l_k + gl_{ik} + e \qquad \text{Eq. (1)}$$

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. (1) was simplified accordingly. Broad-sense heritability (H^2) of agronomic traits was derived from the variance component estimates deriving from Eq. (1) as follows:

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

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 H^2 within locations (*i.e.* repeatability), Eq. (2) was simplified accordingly.

The derivation of PVS BLUPs and H^2 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. (3):

$$y_{ikm} = \mu + g_i + l_k + p_m + gl_{ik} + gp_{im} + pl_{mk} + e \qquad \text{Eq. (3)}$$

Where y_{ikm} is the observed PVS score, and μ, g_i, l_k, and gl_{ik} are as in Eq. (1) 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. (3) was simplified accordingly. $H^2$for PVS traits was derived from the following formula:

$$H^2 = \frac{\sigma_g}{\left(\sigma_g + \frac{\sigma_{gl}}{n_{loc}} + \frac{\sigma_{gm}}{n_{gender}} + \frac{\sigma_e}{n_{rep^*}\, n_{loc^*}\, n_{gender^*}\, n_{farmer}}\right)} \qquad \text{Eq. (4)}$$

In Eq. (4), 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 H^2 (*i.e.* repeatability) by gender and by location, Eq. (4) was simplified accordingly. The 90th percentile of the OA distribution was considered to identify top ranking accessions for men and women. Landraces were benchmarked with the fourth quartile of the distribution of improved lines for all scored traits.

Climatic diversity analyses

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. 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. After aggregating teff georeferenced accessions in Ethiopian administrative regions at the second level (districts), pairwise F_{st}[95]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[96] to check associations between linearized F_{st} (F_{st}/1-F_{st}) and geographic and environmental distances.

A gradient forest (GF) machine-learning approach implemented in R/gradientForest[97,98], 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[99,100] and were calculated with *dbmem()* in R/adespatial[101]. 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. The Euclidean distance between the allelic turnover under the historical and future climates was used to estimate genomic offset, also referred to as genomic vulnerability, to identify areas exposed to teff cultivation depletion pending current teff genetic diversity. 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[63] were used to develop and ensemble projection of rainfall and temperature indices with Climate Data Operators (CDO)[102] 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 and candidate gene analysis

Quantitative trait nucleotides (QTNs) were mapped in a genome wide association study (GWAS). GWAS was performed with R/rMVP[103]using the Fixed and random model Circulating Probability Unification (FarmCPU) method[104] 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[105]. 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[106]. 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[28]. Nucleotide sequences of putative candidate genes were translated into the corresponding proteins and used as queries against Araport11[107]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.

Conflict of Interest Statement

The authors declare no conflict of interest

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, 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.

Acknowledgements

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. 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.

Data Availability

Seed accessions are maintained in Ethiopia at the Ethiopian Biodiversity Institute (EBI) and at the Amhara Regional Agricultural Research Institute (ARARI). Seed requests may be also addressed at the

corresponding author. Raw sequencing data is available at the short-read archive on the NCBI (https://www.ncbi.nlm.nih.gov/sra) with BioProject ID PRJNA758057. Phenotypic data is included in Supplementary Materials of this paper. Genotypic information is available at FigShare (https://figshare.com/) with doi:10.6084/m9.figshare.16451139. Scripts can be found on the corresponding author's GitHub page at https://github.com/mdellh2o/TeffDiversityPanel.

References

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

2. Khoury, C. K. *et al.* Increasing homogeneity in global food supplies and the implications for food security. *Proc. Natl. Acad. Sci.* **111**, 4001–4006 (2014).

3. Jamnadass, R. *et al.* Enhancing {African} orphan crops with genomics. *Nat. Genet.* **52**, 356–360 (2020).

4. Hendre, P. S. *et al.* African {Orphan} {Crops} {Consortium} ({AOCC}): status of developing genomic resources for {African} orphan crops. *Planta* **250**, 989–1003 (2019).

5. Tadele, Z. Orphan crops: their importance and the urgency of improvement. *Planta* **250**, 677–694 (2019).

6. Dawson, I. K. *et al.* 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 (2019).

7. McMullin, S. *et al.* Determining appropriate interventions to mainstream nutritious orphan crops into {African} food systems. *Glob. Food Sec.* **28**, 100465 (2021).

8. Chang, Y. *et al.* The draft genomes of five agriculturally important {African} orphan crops. *Gigascience* **8**, (2019).

9. Liu, H.-J. & Yan, J. Crop genome-wide association study: a harvest of biological relevance. *Plant J.* **97**, 8–18 (2019).

10. Mascher, M. *et al.* Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nat. Genet.* **51**, 1076–1081 (2019).

11. Poland, J. Breeding-assisted genomics. *Curr. Opin. Plant Biol.* **24**, 119–124 (2015).

12. Juliana, P. *et al.* Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. *Nat. Genet.* 1–10 (2019) doi:10.1038/s41588-019-0496-6.

13. Lasky, J. R. *et al.* Genome-environment associations in sorghum landraces predict adaptive traits. *Sci. Adv.* **1**, (2015).

14. Rhoné, B. *et al.* Pearl millet genomic vulnerability to climate change in {West} {Africa} highlights the need for regional collaboration. *Nat. Commun.* **11**, 5274 (2020).

15. Hoffmann, A. A., Weeks, A. R. & Sgrò, C. M. Opportunities and challenges in assessing climate change vulnerability through genomics. *Cell* **184**, 1420–1425 (2021).

16. Aguirre-Liguori, J. A., Ramírez-Barahona, S. & Gaut, B. S. The evolutionary genomics of species' responses to climate change. *Nat. Ecol. Evol. 2021* 1–11 (2021) doi:10.1038/s41559-021-01526-9.

17. Van Etten, J. *et al.* 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 (2019).

18. Fan, S. & Rue, C. The {Role} of {Smallholder} {Farms} in a {Changing} {World}. in *The {Role} of {Smallholder} {Farms} in {Food} and {Nutrition} {Security}* (eds. y Paloma, S., Riesgo, L. & Louhichi, K.) 13–28 (Springer International Publishing, 2020). doi:10.1007/978-3-030-42148-9_2.

19. Terlau, W., Hirsch, D. & Blanke, M. Smallholder farmers as a backbone for the implementation of the {Sustainable} {Development} {Goals}. *Sustain. Dev.* **27**, 523–529 (2019).

20. Ceccarelli, S. & Grando, S. Decentralized-participatory plant breeding: an example of demand driven research. *Euphytica* **155**, 349–360 (2007).

21. van Etten, J. *et al.* Crop variety management for climate adaptation supported by citizen science. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 4194–4199 (2019).

22. Kidane, Y. G. *et al.* Genome wide association study to identify the genetic base of smallholder farmer preferences of durum wheat traits. *Front. Plant Sci.* **8**, 1230 (2017).

23. de Sousa, K. *et al.* Data-driven decentralized breeding increases prediction accuracy in a challenging crop production environment. *Commun. Biol.* (2021) doi:10.1038/s42003-021-02463-

w.

24. Bachewe, F. N. & Taffesse, A. S. Supply response of smallholder households in {Ethiopia}. in *{IFPRI} book chapters* 181–204 (International Food Policy Research Institute (IFPRI), 2018).

25. Seyfu, K. *Eragrostis tef (Zucc.) Trotter: Promoting the conservation and use of underutilized and neglected crops. 12.* (1997).

26. D'Andrea, A. C. T'ef (Eragrostis tef) in ancient agricultural systems of highland Ethiopia. *Econ. Bot.* **62**, 547–566 (2008).

27. Cannarozzi, G. *et al.* Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (Eragrostis tef). *BMC Genomics* **15**, (2014).

28. VanBuren, R. *et al.* Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. *Nat. Commun.* **11**, 1–11 (2020).

29. Cannarozzi, G. *et al.* Technology generation to dissemination: lessons learned from the tef improvement project. *Euphytica* **214**, (2018).

30. Zhu, Q. *et al.* High-throughputdiscoveryofmutationsintefsemidwarfing genes by next-generation sequencing analysis. *Genetics* **192**, 819–829 (2012).

31. Cochrane, L. & Bekele, Y. W. Contextualizing narratives of economic growth and navigating problematic data: Economic trends in Ethiopia (1999–2017). *Economies* **6**, (2018).

32. Girma, D. *et al.* The origins and progress of genomics research on Tef (Eragrostis tef). *Plant Biotechnol. J.* **12**, 534–540 (2014).

33. Assefa, K. *et al.* Genetic diversity in tef [Eragrostis tef (Zucc.) Trotter]. *Front. Plant Sci.* **6**, 1–13 (2015).

34. Woldeyohannes, A. B. *et al.* Current and projected eco-geographic adaptation and phenotypic diversity of Ethiopian teff (Eragrostis teff) across its cultivation range. *Agric. Ecosyst. Environ.* **300**, 107020 (2020).

35. Heffner, E. L., Sorrells, M. E. & Jannink, J. L. Genomic selection for crop improvement. *Crop Science* vol. 49 1–12 (2009).

36. Boeven, P. H. G., Longin, C. F. H. & Würschum, T. A unified framework for hybrid breeding and

the establishment of heterotic groups in wheat. *Theor. Appl. Genet. 2016 1296* **129**, 1231–1245 (2016).

37. Araya, A., Stroosnijder, L., Girmay, G. & Keesstra, S. D. Crop coefficient, yield response to water stress and water productivity of teff (Eragrostis tef (Zucc.). *Agric. Water Manag.* **98**, 775–783 (2011).

38. Casañas, F., Simó, J., Casals, J. & Prohens, J. Toward an Evolved Concept of Landrace. *Front. Plant Sci.* **0**, 145 (2017).

39. Seboka, B. & Deressa, A. Validating farmers' indigenous social networks for local seed supply in central rift valley of {Ethiopia}. *J. Agric. Educ. Ext.* **6**, 245–254 (1999).

40. Delêtre, M., McKey, D. B. & Hodkinson, T. R. Marriage exchanges, seed exchanges, and the dynamics of manioc diversity. *Proc. Natl. Acad. Sci.* **108**, 18249–18254 (2011).

41. Labeyrie, V., Thomas, M., Muthamia, Z. K. & Leclerc, C. Seed exchange networks, ethnicity, and sorghum diversity. *Proc. Natl. Acad. Sci.* **113**, 98–103 (2016).

42. Costanza, S. H., Dewet, J. M. J. & Harlan, J. R. Literature review and numerical taxonomy of Eragrostis tef (T'ef). *Econ. Bot.* **33**, 413–424 (1979).

43. Mancini, C. *et al.* Joining smallholder farmers' traditional knowledge with metric traits to select better varieties of Ethiopian wheat. *Sci. Rep.* **7**, (2017).

44. Fadda, C., Fadda, C., Mengistu, D. K., Kidane, Y. G. & Acqua, M. D. Integrating Conventional and Participatory Crop Improvement for Smallholder Agriculture Using the Seeds for Needs Approach : A Review Integrating Conventional and Participatory Crop Improvement for Smallholder Agriculture Using the Seeds for Needs Approac. *Front. Plant Sci.* (2020) doi:10.3389/fpls.2020.559515.

45. Mengistu, D. K. & Mekonnen, L. S. Integrated Agronomic Crop Managements to Improve Tef Productivity Under Terminal Drought. *Water Stress* (2012) doi:10.5772/30662.

46. Thole, J. M. *et al.* ROOT HAIR DEFECTIVE4 Encodes a Phosphatidylinositol-4-Phosphate Phosphatase Required for Proper Root Hair Development in Arabidopsis thaliana. *Plant Cell* **20**, 381 (2008).

47. Liu, Z., Fan, M., Li, C. & Xu, J. H. Dynamic gene amplification and function diversification of grass-

specific O-methyltransferase gene family. *Genomics* **111**, 687–695 (2019).

48. Held, B. M., Wang, H., John, I., Wurtele, E. S. & Colbert, J. T. An mRNA Putatively Coding for an O-Methyltransferase Accumulates Preferentially in Maize Roots and Is Located Predominantly in the Region of the Endodermis. *Plant Physiol.* **102**, 1001–1008 (1993).

49. McSteen, P. *et al.* 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 (2007).

50. Jia, H. *et al.* A serine/threonine protein kinase encoding gene KERNEL NUMBER PER ROW6 regulates maize grain yield. *Nat. Commun. 2020 111* **11**, 1–11 (2020).

51. S, A. *et al.* Role of Arabidopsis AtPI4Kγ3, a type II phosphoinositide 4-kinase, in abiotic stress responses and floral transition. *Plant Biotechnol. J.* **14**, 215–230 (2016).

52. Wang, K. L.-C., Yoshida, H., Lurin, C. & Ecker, J. R. Regulation of ethylene gas biosynthesis by the Arabidopsis ETO1 protein. *Nat. 2004 4286986* **428**, 945–950 (2004).

53. Valluru, R., Reynolds, M. P., Davies, W. J. & Sukumaran, S. Phenotypic and genome-wide association analysis of spike ethylene in diverse wheat genotypes under heat stress. *New Phytol.* **214**, 271–283 (2017).

54. Iqbal, N. *et al.* Ethylene Role in Plant Growth, Development and Senescence: Interaction with Other Phytohormones. *Front. Plant Sci.* **0**, 475 (2017).

55. Zhu, Q. *et al.* High-Throughput Discovery of Mutations in Tef Semi-Dwarfing Genes by Next-Generation Sequencing Analysis. *Genetics* **192**, 819 (2012).

56. Lemmon, Z. H. *et al.* Rapid improvement of domestication traits in an orphan crop by genome editing. *Nat. Plants* **4**, 766–770 (2018).

57. Sloat, L. L. *et al.* Climate adaptation by crop migration. *Nat. Commun. 2020 111* **11**, 1–9 (2020).

58. Godfray, H. C. J. *et al.* Food Security: The Challenge of Feeding 9 Billion People. *Science (80-.).* **327**, 812–818 (2010).

59. Lee, D. R., Edmeades, S., De Nys, E., McDonald, A. & Janssen, W. Developing local adaptation strategies for climate change in agriculture: A priority-setting approach with application to Latin

America. *Glob. Environ. Chang.* **29**, 78–91 (2014).

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

61. Annicchiarico, P., Russi, L., Romani, M., Pecetti, L. & Nazzicari, N. 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 (2019).

62. Rhoné, B. *et al.* Pearl millet genomic vulnerability to climate change in West Africa highlights the need for regional collaboration. *Nat. Commun.* **11**, 1–9 (2020).

63. Rural, I., Through, L. & Productive, B. *Food Security. In: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*. vol. 544 www.ipcc.ch (2008).

64. Yerima, A. R. I. B. & Achigan-Dako, E. G. 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.* (2021) doi:10.1111/PBR.12930.

65. Abrouk, M. *et al.* Fonio millet genome unlocks {African} orphan crop diversity for agriculture in a changing climate. *Nat. Commun.* **11**, 4488 (2020).

66. Yang, X. *et al.* Early millet use in northern China. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 3726–3730 (2012).

67. Montgomery, J. S. *et al.* Draft Genomes of Amaranthus tuberculatus, Amaranthus hybridus, and Amaranthus palmeri. *Genome Biol. Evol.* **12**, 1988–1993 (2020).

68. Fadda, C. *et al.* Integrating Conventional and Participatory Crop Improvement for Smallholder Agriculture Using the Seeds for Needs Approach: A Review. *Front. Plant Sci.* **11**, 1 (2020).

69. Varshney, R. K. *et al.* Designing {Future} {Crops}: {Genomics}-{Assisted} {Breeding} {Comes} of {Age}. *Trends Plant Sci.* **0**, (2021).

70. Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7**, (2012).

71. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: An analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013).

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

73. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

74. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

75. Danecek, P. *et al.* The variant call format and {VCFtools}. *Bioinformatics* **27**, 2156–2158 (2011).

76. Hijmans, R., Etten, J. Van, … J. C.-R. & 2015, U. Package 'raster'. *slartibardfast.gtlib.gatech.edu*.

77. Reuter, H. I., Nelson, A. & Jarvis, A. An evaluation of void-filling interpolation methods for {SRTM} data. *Int. J. Geogr. Inf. Sci.* **21**, 983–1008 (2007).

78. *Agro-{Ecological} zones of {Ethiopia}.* http://publication.eiar.gov.et:8080/xmlui/bitstream/handle/123456789/2517/AGRO-ECOLOGICALZONES.

79. Database of {Global} {Administrative} {Areas}. (2021).

80. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).

81. Kindt, R. & Coe, R. *Tree diversity analysis; A manual and software for common statistical methods for ecological and biodiversity studies*. *World* (2005).

82. Jones, C. D. *et al.* The {HadGEM2}-{ES} implementation of {CMIP5} centennial simulations. *Geosci. Model Dev.* **4**, 543–570 (2011).

83. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

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

85. Tajima, F. & Nei, M. Estimation of evolutionary distance between nucleotide sequences. *Mol. Biol. Evol.* **1**, 269–285 (1984).

86. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

87. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).

88. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).

89. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, 246 (2011).

90. Garcia-Erill, G. & Albrechtsen, A. Evaluation of model fit of inferred admixture proportions. *Mol. Ecol. Resour.* **20**, 936–949 (2020).

91. Shin, J.-H., Blay, S., McNeney, B. & Graham, J. {LDheatmap}: {An} {R} {Function} for {Graphical} {Display} of {Pairwise} {Linkage} {Disequilibria} {Between} {Single} {Nucleotide} {Polymorphisms}. *J. Stat. Softw.* **16**, 1–9 (2006).

92. Hill, W. G. & Weir, B. S. Variances and covariances of squared linkage disequilibria in finite populations. *Theor. Popul. Biol.* **33**, 54–78 (1988).

93. Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).

94. Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J. & Thompson, R. ASReml User Guide Release 4.1 Functional Specification. *VSN Int. Ltd, Hemel Hempstead, HP1 1ES, UK www.vsni.co.uk* (2014).

95. Weir, B. S. & Cockerham, C. C. Estimating F-Statistics for the Analysis of Population Structure. *Evolution (N. Y).* **38**, 1358 (1984).

96. Dray, S. & Dufour, A.-B. The ade4 Package: Implementing the Duality Diagram for Ecologists. *J. Stat. Softw.* **22**, 1–20 (2007).

97. Ellis, N., Smith, S. J. & Pitcher, C. R. Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93**, 156–168 (2012).

98. Fitzpatrick, M. C. & Keller, S. R. Ecological genomics meets community-level modelling of

biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecol. Lett.* **18**, 1–16 (2015).

99. Dray, S., Legendre, P. & Peres-Neto, P. R. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices ({PCNM}). *Ecol. Modell.* **196**, 483–493 (2006).

100. Griffith, D. A. & Peres-Neto, P. R. Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology* **87**, 2603–2613 (2006).

101. Stéphane Dray, A. *et al.* Package 'adespatial' Title Multivariate Multiscale Spatial Analysis. (2021) doi:10.1890/11-1183.1.

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

103. Yin, L. *et al.* {rMVP}: {A} {Memory}-efficient, {Visualization}-enhanced, and {Parallel}-accelerated tool for {Genome}-{Wide} {Association} {Study}. *bioRxiv* 2020.08.20.258491 (2020) doi:10.1101/2020.08.20.258491.

104. Liu, X., Huang, M., Fan, B., Buckler, E. S. & Zhang, Z. Iterative {Usage} of {Fixed} and {Random} {Effect} {Models} for {Powerful} and {Efficient} {Genome}-{Wide} {Association} {Studies}. *PLOS Genet.* **12**, e1005767 (2016).

105. VanRaden, P. M. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* **91**, 4414–4423 (2008).

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

107. Cheng, C.-Y. *et al.* Araport11: a complete reannotation of the {Arabidopsis} thaliana reference genome. *Plant J. Cell Mol. Biol.* **89**, 789–804 (2017).