Core ideas

- 8,778 SNPs and 13 agronomic traits characterized a panel of 423 finger millet landraces.
- 4 clusters of accessions coincided with major geographic areas of finger millet cultivation.
- A comparison of phenotypic and genomic data indicated a complex diversification history.
- This was confirmed by the analysis of allotetraploid finger millet's separate sub-genomes.
- Comprehensive new knowledge for intra- and inter-regional breeding is provided.

Genomic and phenotypic characterization of finger millet indicates a complex diversification history

Jon Bančič[1,2,*], Damaris A. Odeny[3], Henry F. Ojulong[3], Samuel M. Josiah[4], Jaap Buntjer[1], R. Chris Gaynor[1], Stephen P. Hoad[2], Gregor Gorjanc[1], Ian K. Dawson[2]

[1] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush Research Centre, Midlothian EH25 9RG, UK

[2] Scotland's Rural College (SRUC), Kings Buildings, West Mains Road, Edinburgh EH9 3JG, UK

[3] International Crops Research Institute for the Semi-Arid Tropics, ICRAF House, UN Avenue, Gigiri, P.O. BOX 39063-00623 Nairobi, Kenya

[4] University of Georgia, Department of Horticulture, Athens, GA, 30602-7273, USA

Abbreviations: AMOVA, analysis of molecular variance; CV, coefficient of variation; DAPC, Discriminant Analysis of Principal Components; DArTseq, Diversity Arrays Technology sequencing; GRM, genomic relationship matrix; GWAS, genome-wide association study; LD, linkage disequilibrium; MAF, minor allele frequency; MLM, mixed linear model; MTA, marker-trait association; NJ, neighbour-joining; PCA, principal component analysis; RR-BLUP, ridge regression-best linear unbiased genomic prediction; SNP, single nucleotide polymorphism

ABSTRACT

Advances in sequencing technologies mean that insights into crop diversification aiding future breeding can now be explored in crops beyond major staples. For the first time, we use a genome assembly of finger millet, an allotetraploid orphan crop, to analyze DArTseq single nucleotide polymorphisms (SNPs) at the sub-genome level. A set of 8,778 SNPs and 13 agronomic traits characterizing a broad panel of 423 landrace accessions from Africa and Asia suggested the crop has undergone complex, context-specific diversification consistent with a long domestication history. Both Principal Component Analysis and Discriminant Analysis of Principal Components of SNPs indicated four groups of accessions that coincided with the principal geographic areas of finger millet cultivation. East Africa, the considered origin of the crop, appeared the least genetically diverse. A Principal Component Analysis of phenotypic data also indicated clear geographic differentiation, but different relationships among geographic areas than genomic data. Neighbour-joining trees of sub-genomes A and B showed different features which further supported the crop's complex evolutionary history. Our genome-wide association study indicated only a small number of significant marker-trait associations. We applied then clustering to marker effects from a ridge regression model for each trait which revealed two clusters of different trait complexity, with days to flowering and threshing percentage among simple traits, and finger length and grain yield among more complex traits. Our study provides comprehensive new knowledge on the distribution of genomic and phenotypic variation in finger millet, supporting future breeding intra- and inter-regionally across its major cultivation range.

INTRODUCTION

Diversifying crop production is an important global objective to address human and environmental health concerns, such as malnutrition and the use of intensive unsustainable monoculture production systems (Bančič et al., 2021; von Grebmer et al., 2014). To achieve diversification, we need to increase focus on under-researched crops, also known as orphan crops. Some of these orphan crops are rich in micro- and macro-nutrients and can complement other crops in food production systems, even when farming conditions are adverse (Kamenya et al., 2021; Mustafa et al., 2019). Apart from their intrinsic value, many orphan crops have complex demographic histories that can shed light on broader crop domestication and diversification processes (Meyer & Purugganan, 2013). The high cost of generating genomic resources has however traditionally prevented genomic analyses for orphan crops, but a significant cost-reduction in high-throughput sequencing in the last decade has provided new opportunities for their research (Jamnadass et al., 2020).

Finger millet (*Eleusine coracana* (L.) Gaertn. subsp. *coracana*; Poaceae, subfamily Chloridoideae) is an annual small-grained cereal and an orphan crop with an essential role in smallholder food production systems in parts of Africa and Asia. The crop's attractive characteristics include its rich nutritional profile, versatility in food usage, good storage properties, high market value, adaptability to poor production conditions, and flexibility in integrating into various farming approaches (Odeny et al., 2020; Sood et al., 2019). According to de Wet et al. (1984), the allotetraploid crop (*2n=4x*=36; genome constitution AABB, reported disomic inheritance) is thought to have been domesticated from wild *E. coracana* subsp. *africana* in either Uganda or the Ethiopian highlands of East Africa around five millennia ago. The domesticated subsp. *coracana* then spread to southern Africa while broadly maintaining sympatry with the subsp. *africana* wild

form, and remaining in proximity to other *Eleusine* species (see a distribution of the eight *Eleusine* species in Africa depicted in **Fig. 1**). According to de Wet et al. (1984), the crop was introduced to India around three millennia ago, from where it was then dispersed further, including to Nepal. Therefore, East Africa is considered the primary centre of diversity and India an important secondary centre. The immediate wild progenitor of finger millet is believed to have arisen in East Africa by hybridization between two diploid species, the first of which being *E. indica*, which still occurs widely across Africa and is considered the (maternal) donor of what has become finger millet's A sub-genome; and the second, still unknown (and now possibly extinct), diploid pre-B sub-genome donor (Liu et al., 2014; Zhang et al., 2019).

In Africa and Asia, subsistence farmers who rely on finger millet mostly grow landrace varieties, and systematic genetic improvement has been limited. This and its complicated biology reflect finger millet's status as an orphan crop. The crop predominantly self-pollinates and has small flowers, which challenges artificial crosses (Dida & Devos, 2006). Moreover, its genomic status as an allotetraploid with an unknown B sub-genome donor has hampered the development of genomic resources. This is because difficulties in distinguishing between sub-genomes and thus the inaccurate calling of homologous versus homeologous single nucleotide polymorphism (SNP) positions are a possibility (Hatakeyama et al., 2018; Hittalmani et al., 2017). Landrace varieties however provide important opportunities to explore crop domestication and diversification. Many have been sampled for conservation in genebanks (Upadhyaya et al., 2006) and these accessions are available for use in breeding programs. They are also a resource to explore the crop's diversification history, a topic that has so far received only limited attention.

The recent availability of a draft genome sequence and a robust linkage map for finger millet transform the potential for using genomic information to assist breeding and further understand

the crop's diversification history (Odeny et al., 2020). This paper uses a combination of genomic and phenotypic data to explore a broad panel of 423 finger millet landrace accessions sampled across its main cultivation regions in Africa and Asia. We make use of Diversity Arrays Technology sequencing (DArTseq) (Sansaloni et al., 2011) SNP markers that are, for the first time, placed onto finger millet sub-genomes using the recent genome assembly (https://phytozome-next.jgi.doe.gov/info/Ecoracana_v1_1). Our objective is to characterize the crop's genomic and phenotypic variation to explore the diversification process and to provide insights for future breeding across its main cultivation range. We here explore multiple features of finger millet variation, including the geographic structuring of genomic and phenotypic diversity, sub-genome specific diversity profiles, germplasm migration events amongst geographic areas, and genetic architecture and selection patterns for agronomic traits. Our analysis provides essential information for the future development of finger millet and is a model for exploring other orphan crops.

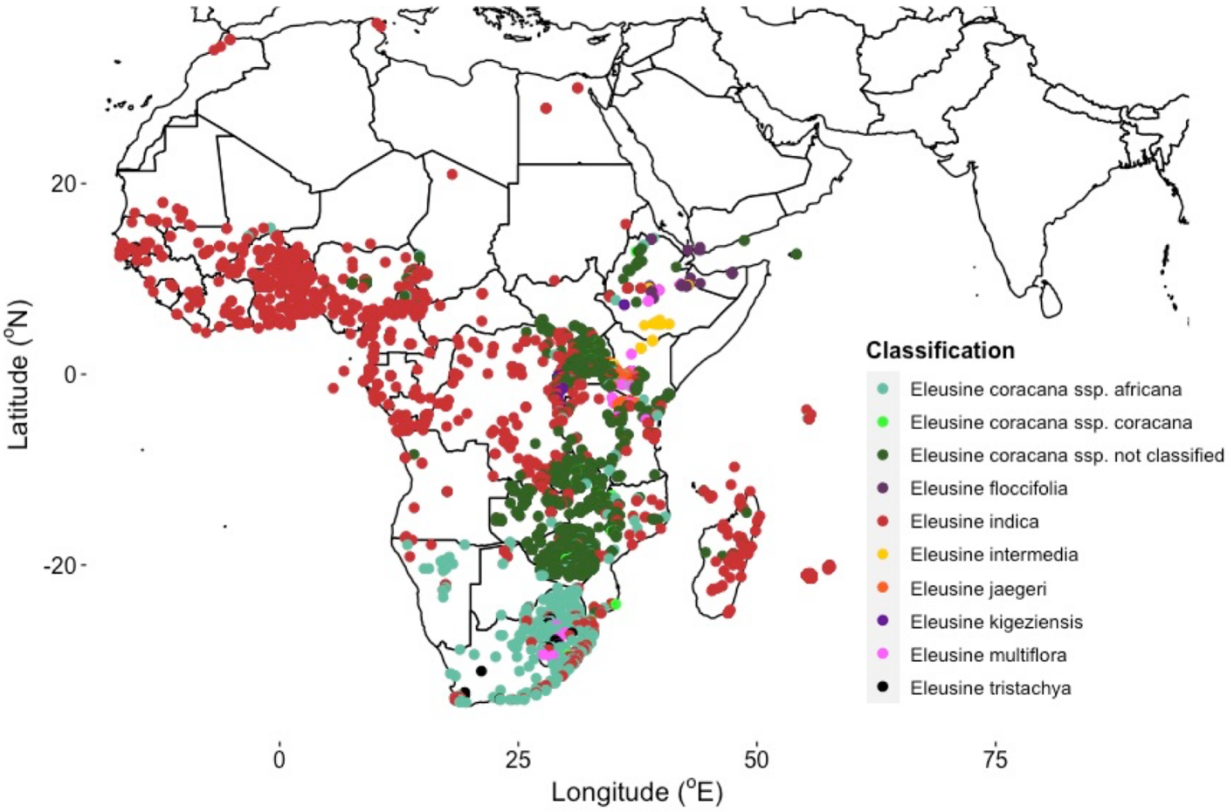


Fig. 1: Distribution of eight *Eleusine* species across the African continent. A total of 9,250 known locations of *Eleusine* were extracted from the Global Biodiversity Information Facility (GBIF; https://www.gbif.org/). Some *Eleusine coracana* entries in GBIF were not classified at the subspecies level ('not classified' in key).

MATERIALS AND METHODS

Plant material

This study uses the most extensive set of jointly genotyped and phenotyped finger millet landrace accessions to date from the crop's main cultivation regions in Africa and South Asia. The panel initially contained 458 accessions (later reduced to 423 accessions for analysis) that as well as encompassing the main cultivation regions included a small number of accessions collected more widely (**Tab. S1**). In total, 19 accessions (designated as "other") were collected outside the main cultivation regions or are of unknown origin. Our panel is from the ICRISAT genebank's Core

Collection of finger millet assembled by Upadhyaya et al. (2006). The 458 accessions are a subset with extensive phenotypic variation and have been the focus of breeder's activities in recent years. Most of the Core Collection was initially sampled directly from farmers' fields, although sometimes accessions were sampled from local markets (see, e.g., Rao, 1980).

Collecting and processing genomic data

Leaf tissue was taken from single individuals of each of the 458 accessions, grown in a greenhouse at ICRISAT in Nairobi and dried with silica gel. Genomic DNA (gDNA) was extracted from finely-ground leaf material using the ISOLATE II Genomic DNA Kit (Bioline Pty Ltd) and according to the manufacturer's instructions. The purity and quantity of extracted gDNA was determined by gel electrophoresis and a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA), respectively, with a final dilution of gDNA to 50 ng/µl. Genomic DNA was then delivered to the Integrated Genotyping Service and Support (IGSS) facility at the Bioscience eastern and central Africa–International Livestock Research Institute (BecA–ILRI) hub in Nairobi. This was for library construction and DArTseq data generation with SNP positions' assignment using the v1.1 finger millet genome assembly (https://phytozome-next.jgi.doe.gov/info/Ecoracana_v1_1), and initial output quality control steps, using methods described previously (Sansaloni et al., 2011).

The IGSS facility generated 70,906 raw SNPs via DArTseq, which we then quality filtered using TASSEL (version 5.0; Bradbury et al. 2007) as illustrated in **Fig. S1**. The accessions and SNPs were filtered using a minimum call rate of 70%, which reduced our initial sample size of 458 accessions to 423 (**Tab. S1**). This constituted our final accession set for later data analyses. We then retained only those SNPs with a minor allele frequency (MAF) > 0.01. In further screening,

we removed SNPs with a heterozygosity level greater than $2pq \times (1 - F)$, where p and q are the frequencies of the two allele states, and F is the inbreeding coefficient, for which a value of 0.5 was chosen due to the self-pollinating nature of finger millet. This last filter was applied to remove SNP calls that most likely originated from incorrectly collapsing homeologous positions across the two sub-genomes.

Measuring linkage disequilibrium decay along chromosomes

Since our analysis is the first to use SNP marker positions assigned using a finger millet genome assembly, we initially explored associations among our high-quality SNP data set in our specific germplasm panel. We calculated linkage disequilibrium (LD) decay (r^2) between all pairwise intra-chromosomal SNP combinations using a full correlation matrix either uncorrected or corrected for bias based on underlying genetic structure across the accessions. Calculations of r^2 were performed in the R package *LDcorSV* (Mangin et al., 2012). Pairwise r^2 values were then plotted against chromosomal physical distance. The decay curve was fitted based on Hill and Weir (1988) using R code from Marroni et al. (2011) and was then used to estimate the distance at which r^2 decreased to 0.2. We estimated LD decay for each chromosome separately, for each sub-genome and for the genome as a whole.

Genetic structure, differentiation and diversity

Genome-wide genetic structure

We used four approaches to characterize and visualize genetic structure. First, we used R packages *ade4* (Dray & Dufour, 2007) and *adegenet* (Jombart, 2008; the *find.clusters* function) to undertake Principal Component Analysis (PCA) and Discriminant Analysis of Principal Components (DAPC). For the latter, the optimum PC number and cluster number (*k*) were set based on the change in the curve shape of profiles (**Figs. S2**, **S3**). Second, we constructed a genomic relationship matrix (GRM) using a centred-identity-by-state method (Endelman & Jannink, 2012) in TASSEL. A heatmap of the GRM, calculated using the unweighted pair group method with arithmetic mean (UPGMA) cluster algorithm, was visualized using the *heatmap* R function (R Core Team, 2019). Third, we constructed using TASSEL and plotted with R package *phangorn* (Schliep, 2011) an unweighted neighbour-joining (NJ) tree (Saitou & Nei, 1987) of accessions. We extended NJ analysis to consider not only genome-wide markers but SNPs pooled at the sub-genome level. Fourth, we visualized the country-level geographic distribution of finger millet accessions assigned to their genetic clusters using R packages *rworldmap* (South, 2011) and *ggplot2* (Wickham, 2009).

Detection of introgression and gene flow

We used Treemix (Pickrell & Pritchard, 2012) to infer the most likely evolutionary history and evidence for introgression amongst groups of accessions assigned to four geographic areas of sampling that we term ‘East Africa’, ‘Southern Africa’, ‘India’ and ‘Nepal’ (**Tab. S1**). For simplification, the small number of accessions sampled from countries outside these locations (9 of all 404’ known’-location accessions) were assigned to the most proximate of our defined four areas (indicated with ellipses in **Fig. 2d**). The 19 accessions designated as “other” (see Plant

material section) were excluded from geographic area analysis. **Table S1** provides complete information, but, in brief: the area 'East Africa' included a small number of additional accessions (N = 5) from Nigeria and Senegal; and 'India' included a few extra lines (N = 4) from Sri Lanka, the Maldives and Pakistan. Using Treemix, maximum-likelihood population trees were constructed based on genome-wide SNPs using blocks of 50 SNPs and 'East Africa' rooted as an out-group. The number of tested migration events was varied from zero to three. Bootstrap replicates were generated using 50 SNPs to evaluate the robustness of tree topology, and Treemix R plotting functions were used to visualize results.

Chromosome-level nucleotide diversity and differentiation

We analyzed diversity and differentiation at a chromosome level for both sub-genomes of finger millet for accessions assigned to four geographic areas of sampling (areas as explained above). For each area, we calculated $\boldsymbol{\pi}$ (Nei and Li, 1979) as our estimator of diversity and pairwise F_{ST} values (Weir and Cockerham, 1984) as our estimator of differentiation, using VCFtools (Danecek et al., 2011) and a 1 Mb non-overlapping sliding window. Genetic differentiation estimates were calculated for all six possible pairwise combinations of geographic areas. Results were plotted against chromosomal positions using R package *ggplot2*.

Summarized gene diversity and differentiation statistics

We summarised genome-wide gene diversity and differentiation for the four geographic areas. Genome-wide gene diversity (H; Nei, 1973) and pairwise F_{ST} values based on all high-quality SNPs were calculated with R package *hierfstat* (Goudet, 2005). We also computed H values for the four defined genetic clusters (see above) of finger millet accessions (these approximate our four defined geographic areas, as will be explained below). We further extended our analysis of

geographic areas to consider genome-wide markers and diversity at the sub-genome level. We also used R package *pegas* (Paradis, 2010) for an analysis of molecular variance (AMOVA) that partitioned genetic variation within and among our geographic areas (or clusters) as part of the total panel, based on 100,000 permutations.

Collecting and analyzing phenotypic data

To collect information on phenotypic variation in finger millet, we characterized our initial panel of 458 accessions for 13 life history and other traits (**Tab. S2**) in a field trial that used a complete randomized block design with two replications. The trial was conducted over the long rainy season of 2015 at the Kenya Agricultural and Livestock Research Organization field station at Kiboko in Eastern Kenya (coordinates: 2° 20' N, 37° 45' E; altitude: 960m; annual temperatures: min. 16.6°C, max. 29.4°C, average 23.0°; sandy clay loam calcareous soil). Seeds of each accession were drilled in 2 m long plots of two rows spaced 50 cm apart. Within rows, plants were thinned after establishment to a spacing of 10 cm. Five plants for each plot were randomly selected for data recording, though some traits were recorded on a whole plot basis (**Tab. S2**). The traits we measured included features of morphology, physiology and yield that are important for crop production and the integration of finger millet in mixed crop farming systems (Dawson et al., 2019), including in intercrop systems (Brooker et al., 2015), that are of particular interest in breeding research (Bančič et al., 2021).

Our analysis of variance of phenotypic data used the following statistical model:

$$y_{ikn} = \mu + r_k + g_i + b_{kn} + e_{ikn},$$

where y_{ikn} is generally the average phenotype of five plants of the i^{th} accession tested in the k^{th} replication (k = 1, 2) and in the n^{th} block (n = 1,…, n); μ is the intercept; r_k is the fixed effect of a replication; g_i is the random accession (genotype) effect assuming $g_i \sim N(0, \sigma_G^2)$; b_{kn} is the random block effect assuming $b_{kn} \sim N(0, \sigma_b^2)$; and e_{ikn} is the random residual assuming $e_{ikn} \sim N(0, \sigma_e^2)$. Here $N(.,.)$ denotes a normal random variable and σ_G^2, σ_b^2 and σ_e^2 are respectively variances between accessions, blocks and residuals. We treated accessions as fixed effects to obtain best linear unbiased estimates (BLUEs) for each phenotypic trait. Models were fitted using R package *ASReml-R* (version 4.1.0.90; Butler et al. 2017). We estimated broad-sense heritability across the two replications of the trial (p = 2) for each trait as $H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_e^2}{p}}$. Distributions of BLUEs for each phenotypic trait and Pearson's correlation coefficients (r ; Pearson, 1895) between all pairwise combinations were calculated and visualized using R (R Core Team, 2019). The coefficient of variation (CV) for each phenotypic trait was calculated as $(\sigma/\bar{x}) \times 100\%$ using BLUEs.

We then performed Principal Component Analysis using BLUEs of traits to understand the phenotypic structure in these data using R package *factoextra* (Kassambara & Mundt, 2020). Variation of BLUEs within our defined geographic areas and genetic clusters was also visualized using boxplots.

Trait genetic architecture and further genomic-phenotypic comparison

Genome-wide association analysis

To test each SNP's effect on the 13 phenotypic traits, we ran a genome-wide association study (GWAS) in TASSEL using a linear mixed model that corrected for genetic structure and cryptic relatedness (Yu et al., 2006). BLUEs calculated for each trait were taken as phenotype, and 8,778 SNPs were taken as genotype. The model estimated variance components only once using the P3D method. A Bonferroni threshold of 5% was applied to account for multiple testing to declare significant marker-trait associations (MTAs). Manhattan and quantile-quantile (Q-Q) plots of the GWAS results were visualized with R package *qqman* (Turner, 2014). Putative candidate genes were selected using a 25 kb genomic interval both upstream and downstream of a significant MTA. The interval was queried against the current finger millet genome assembly using the Integrated Genomics Viewer (IGV, v. 2.82; Robinson et al., 2011).

Clustering SNP effects

To further evaluate the genetic architecture of the 13 phenotypic traits, we explored the distributions of SNP effects from a ridge regression model (RR-BLUP) (Kooke et al., 2016). BLUEs calculated for each trait scaled by their standard deviation were taken as phenotype, and an imputed 8,778 SNP matrix were taken as genotype. A *k*-nearest neighbour imputation of the SNP matrix was performed in TASSEL using default parameters and RR-BLUP models were fitted in R package *AlphaMME* (https://github.com/gaynorr/AlphaMME). SNP effects from the RR-BLUP model were then used to: first, calculate pairwise Euclidean distances over the first five mathematical moments in R package *moments* (v0.14; Komsta and Novomestky, 2011); and, second, construct a dendrogram from the Euclidean distance matrix using the UPGMA cluster algorithm with the *hclust* R function (R Core Team, 2019).

Comparison of phenotypic and genome-wide gene diversity by geographic area

A simple yet useful way to shed light on the particular selection histories of crops as they take different diversification pathways within specific geographic contexts is to compare phenotypic diversity levels with levels of underlying genome-wide gene diversity. Here, we adopt a straightforward approach for initial comparisons that involves individual phenotypic trait CV values and gene diversity (H) values for genome-wide SNP data. We calculate CV/H as the comparator for each of our four defined geographic areas to check primarily for rank differences across areas that may indicate particular phenotypic selection pressures by area. This type of approach is exemplified classically in studies of the diversification of tomato, where high phenotypic variation observed at specific traits is accompanied by overall underlying genomic diversity bottlenecks (Rodríguez et al., 2011).

RESULTS

Genome-wide SNP data cover the entire finger millet chromosome complement but occur at a higher density in the A sub-genome

The IGSS facility generated 70,906 raw SNPs, which were filtered into a final marker set of 8,778 SNPs for 423 accessions for subsequent analyses. Eight thousand and ninety-six (8,096) SNPs out of the 8,778 had been previously mapped to chromosomes (**Tab. S3**). Initial analysis showed that the full complement of finger millet chromosomes was covered but that chromosome-level SNP density was higher for the A sub-genome (**Tab. 1**; approx. twice the density of markers of the B sub-genome). Across both sub-genomes, the mean SNP density was ~8.6 per Mb. Consistent with expectations of significantly lower recombination in central chromosomic regions of selfing cereals (e.g., Bustos-Korts et al., 2019); SNP density was generally considerably higher toward the ends of chromosomes (**Figs. S4**, **S9c**). The proportion of markers removed from our initial raw

data due to likely being homeologous was a relatively high 3.9% (2,733 out of 70,906; markers above the red curve in **Fig. S1e**). Therefore, precautions with homeologous markers are indicated for finger millet. The observed heterozygosity for the markers in our final set of 8,778 SNPs was very low (median 0.048 and mean 0.057; **Tab. S3**), indicating that most finger millet accessions are highly inbred.

Linkage disequilibrium decay along chromosomes shows expected patterns for both sub-genomes

Our work is the first that has used a finger millet genome assembly to assign physical SNP positions for a broad accession panel, so understanding patterns of LD in A and B sub-genomes is of particular relevance. Our calculations indicated the expected overall pattern of LD decay and the importance of correcting for underlying genetic structure among accessions. For the genome as a whole, r^2 decayed to 0.2 at a distance of 106 kb when correcting for genetic structure (compared to a distance of 1.77 Mb for the naïve model). Decay was slower for the B sub-genome ($r^2 = 0.2$ at 168 kb, with correction) than the A sub-genome ($r^2 = 0.2$ at 88 kb), and varied markedly across chromosomes within sub-genomes (**Fig. S5**). Overall, the relatively slow rate of LD decay was consistent with other self-pollinating crops (Flint-Garcia et al., 2003), including other millets (Jaiswal et al., 2019).

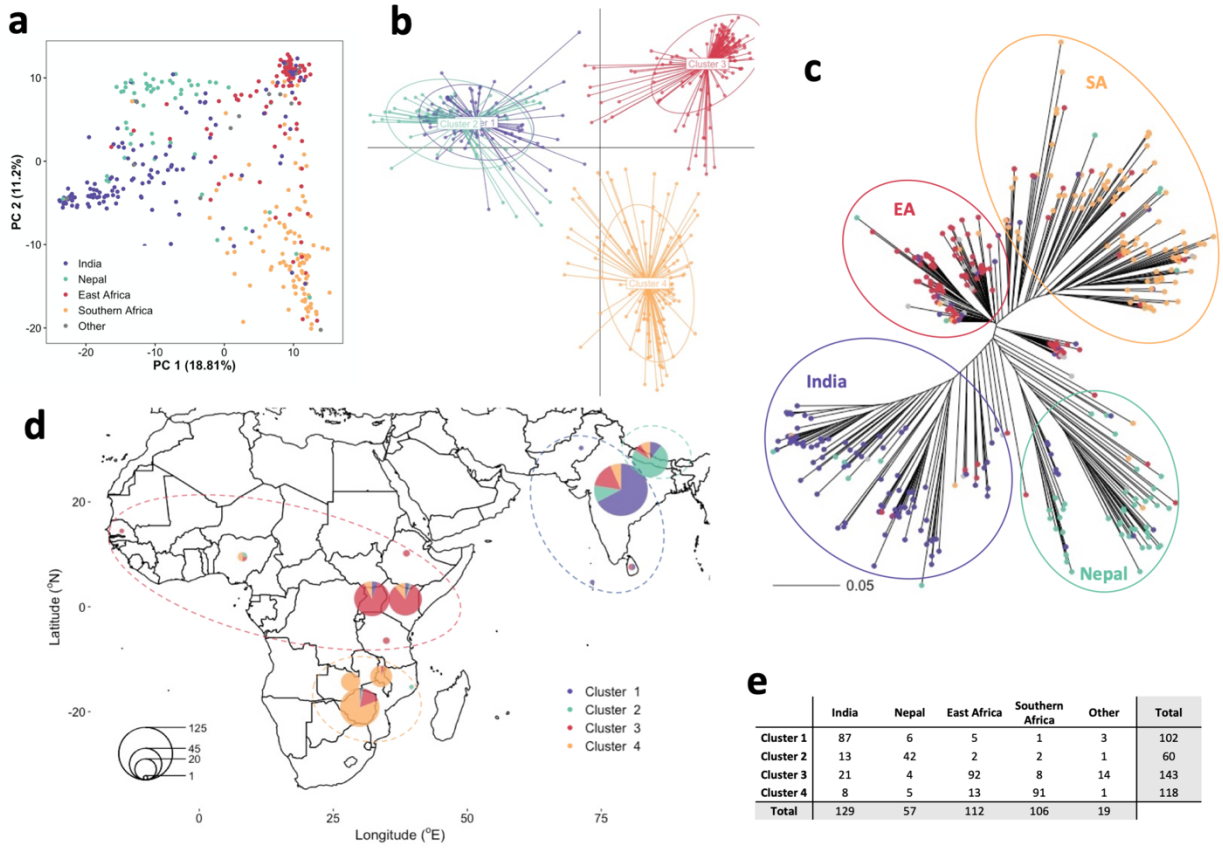
Tab. 1. Distribution of SNPs by finger millet chromosomes. SNP density per chromosome was calculated for individual 1 Mb non-overlapping windows as shown in **Figs. S4** and **S9c**.

Chromosome	Number of SNPs	Chromosome length (Mb)	SNP density (per Mb)
1A	742	57.13	12.4
2A	683	60.90	11.1
3A	570	53.87	10.9
4A	618	39.68	15.4
5A	1057	66.30	15.4
6A	357	59.47	5.9
7A	553	48.45	11.2
8A	307	47.22	7.6
9A	486	45.36	11.6
Sub-genome A	5,373	53.15	11.3
1B	196	72.49	3.7
2B	323	70.83	5.3
3B	136	63.91	2.9
4B	158	50.88	2.9
5B	659	80.22	7.8
6B	244	74.22	5.0
7B	267	58.85	7.1
8B	294	60.62	5.4
9B	446	62.05	7.4
Sub-genome B	2,723	65.20	5.7
Scaffolds	682		
Whole genome	8,778	1072.44	8.6

Genome-wide genetic structure reveals strong geographic differentiation

Our analysis of genome-wide genetic structure in finger millet revealed clear differentiation patterns (**Fig. 2**). Both PCA and DAPC analyses (DAPC using an optimum principal component number PC = 4 according to **Fig. S2** and an optimal cluster number k = 4 according to **Fig. S3**) revealed a genetic structure that corresponded with the four sampled geographic areas of finger millet cultivation (**Fig. 2a, b, d**). Accessions from Africa and Asia regions were separated along the first PC, which explained 18.8% of the total variation, while the second PC, which explained

11.2% of the total variation, discriminated between geographic areas within the two continents (**Fig. 2a**). The overlap between genetic clusters and geographic areas was particularly strong in Africa (**Fig. 2e**, **Tab. S4**). The clear overall geographic structuring of genetic variation, with a degree of admixture, was also evident in our NJ tree (**Fig. 2c**), on the map of sampled geographic areas showing accessions' assignments to genetic cluster groups (**Fig. 2d**), and in a heatmap of GRM with UPGMA clustering (**Fig. S6**). The more extensive admixture of finger millet in Asia than Africa, most clearly observed in **Fig. 2d**, is consistent with more formal breeding of the crop in Asia, supported by cross-regional germplasm transfer. For example, in India, an improvement program initiated in the 1960s involved crosses between African and Asian accessions, resulting in 'Indaf' varieties (Mirza & Marla, 2019; see more in next section). In further NJ analysis that considered SNPs for sub-genomes A and B separately, clear geographic structuring of genetic variation was evident in both cases (**Fig. S7**).



e

	India	Nepal	East Africa	Southern Africa	Other	Total
Cluster 1	87	6	5	1	3	102
Cluster 2	13	42	2	2	1	60
Cluster 3	21	4	92	8	14	143
Cluster 4	8	5	13	91	1	118
Total	129	57	112	106	19	

Fig. 2: Geographical distribution and genetic structure of 423 finger millet accessions. a, Score plot using the first two principal components (PCs) illustrates the genetic structure of finger millet coloured according to geographic areas. **b**, Genetic clusters identified by DAPC and coloured according to their corresponding geographic area in **d**). **c**, A phylogenetic neighbour-joining tree with tips coloured by geographic area and enclosing circles representing generalized clades. **d**, Geographical distribution of accessions with pie chart size representing the number of accessions collected in a particular country and slice size representing the probability of belonging to a specific genetic cluster. Dotted ellipses indicate the extent of our applied geographic areas (drawing in small groups of outliers; see Materials and Methods and **Tab. S1** for further information on geographic area assignments). **e**, A table of accessions assigned to geographic areas and genetic clusters; accessions without known location are designated as 'Other'.

Genetic introgression and gene flow analysis suggests a migration event between East Africa and India

Our Treemix analysis of past admixture events in finger millet suggested a potential historic introgression from East Africa to India. Using 'East Africa' as an out-group, the maximum-likelihood tree without migration events (**Figs. 3a** and **S8a**), accounting for drift alone, corresponded to the phylogenetic tree (see **Fig. 2c**). When migration events were allowed, a single event from 'East Africa' to 'India' was inferred (**Figs. 3b** and **S8b**). These results are consistent with the cultivated finger millet's demographic history as previously speculated by non-genomic analysis approaches (de Wet et al., 1984; Hilu & de Wet, 1976). These findings are also consistent with known breeding introductions from Africa to Asia (see previous section).

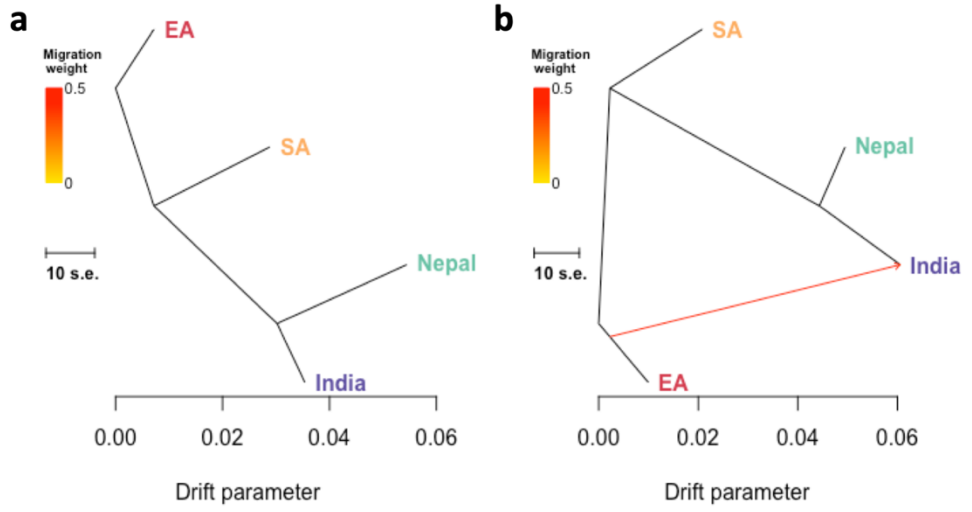


Fig. 3: Inferred finger millet maximum-likelihood population trees with admixture events. The structure of the graphs inferred by Treemix for four geographic areas of finger millet. **a**, with no migration events. **b**, with one migration event. The migration arrow is coloured according to its weight and represents the fraction of ancestry derived from the migration edge. Horizontal branch lengths are proportional to the amount of genetic drift that has occurred on the branch. The scale bar shows ten times the average standard error (s.e.) of the accessions in the sample. The residual fit from the graph is shown in **Fig. S8**.

Chromosome-level nucleotide diversity and differentiation profiles vary by region

Our analysis of nucleotide diversity (π) along chromosomes showed differences in relative diversity for geographic areas by chromosome, including for homeologous (A and B sub-genome) chromosomes (**Fig. S9a)**. Notable was relatively high diversity along chromosome 1A for 'Nepal' (not seen on chromosome 1B) and along chromosomes 5B and 9B for 'Southern Africa' (not seen on chromosomes 5A and 9A). Pairwise chromosome-level F_{ST} values for geographic areas (**Fig. S9b**) reflected these different diversity profiles, again indicating chromosome-specific, sub-genome-based differences. Of note was high differentiation between both of our African geographic areas and both our Asian geographic areas along chromosome 5A that was not replicated on chromosome 5B.

Summarized gene diversity statistics confirm regional differentiation and sub-genomic differences

Our calculation of genome-wide gene diversity (H) values indicated that the 'Nepal' region contained the most diversity and the 'East Africa' region the least (**Tab. 2**). The ranking of diversity levels corresponded when the analysis was repeated based on genetic clusters that approximate geographic areas. Estimates were, as expected, overall lower when based on genetic clusters, as the most prominent within-area genetic admixture has in this case been reassigned (see **Fig. 2**).

379 Pairwise F_{ST} values summarised for genome-wide SNPs (**Tab. S5**) confirmed our PCA and DAPC
380 results (**Fig. 2a, b, d**), which revealed primary geography-based partitioning between Africa and
381 Asia and then between areas within these regions. The highest differentiation was between
382 'Southern Africa' and 'Nepal' accessions (F_{ST} = 0.167). Our two AMOVA analyses revealed that
383 16% of total genome-wide variation partitioned among our four geographic areas and 24% among
384 our corresponding but genetically-defined clusters ($p \leq 0.001$ that no structuring in both cases).

385 In the case of gene diversity (H) calculations, we also analyzed geographic areas for separate A
386 and B sub-genomes (**Tab. 2**). The ranking of diversity by geographic area varied for the two sub-
387 genomes, with 'Nepal' (still, compared to the entire genome) ranking highest for the A sub-
388 genome but 'Southern Africa' highest for the B sub-genome; 'East Africa' consistently ranked
389 lowest. The observed diversity ranking change is reflected in the relative spread of accessions from
390 each geographic area in sub-genome NJ trees (**Fig. S7**). It appears to be based on changes in
391 relative diversity for specific pairs of homeologous chromosomes (**Fig. S9a**), notably
392 chromosomes 5B versus 5A, and 9B versus 9A (diversity relatively high at 5B and 9B for
393 'Southern Africa'; see also above).

394

Tab. 2. Characterization of gene diversity (H). Genome-wide measurement of finger millet gene diversity was undertaken for four geographic areas (left value) and corresponding genetic clusters identified by DAPC analysis (right value). The values presented for individual sub-genomes are for geographic areas only. The total sample size for genome-wide estimates varies because of the non-assignment of some accessions to geographic areas (see Materials and Methods). The area or cluster with highest diversity is underlined.

Geographic area / Genetic cluster	Sample size	H		
		Genome-wide	Sub-genome A	Sub-genome B
India / Cluster 1	129 / 102	0.259 / 0.215	0.293	0.212
Nepal / Cluster 2	57 / 60	<u>0.282</u> / <u>0.265</u>	<u>0.336</u>	0.252
EA / Cluster 3	112 / 143	0.199 / 0.179	0.200	0.208
SA / Cluster 4	106 / 118	0.260 / 0.252	0.230	<u>0.368</u>

Significant variation in phenotypic traits partitions by region

Our analysis of 13 phenotypic traits for 423 finger millet accessions revealed extensive variation (CV from 8.58 for threshing percentage to 48.25 for productive tiller number) and medium-to-high H^2 values (from 0.35 for grain yield to 0.95 for days to flowering). The level of variation detected and heritability values were generally consistent with the crop's previous field trials (e.g., Bharathi, 2011; Manyasa et al., 2016).

Summary statistics are presented in **Tab. S6**, and trait distributions and between-trait correlations in **Fig. S10**. The strongest positive correlation was between leaf length and plant height (r = 0.83, $p < 0.001$), and the strongest negative correlation between leaf width and the number of productive tillers (r = -0.64, $p < 0.001$). Considering the three key traits of grain yield, plant height and days to flowering, a medium-level positive correlation was observed between plant height and days to flowering (r = 0.50, $p < 0.001$), a moderate positive correlation between plant height and grain yield (r = 0.19, $p < 0.001$), and no correlation between days to flowering and grain yield ($r < 0.01$,

NS). Principal component analysis using BLUEs of traits further illustrated the levels of correlation between different traits (**Fig. 4a**).

Similar to genomic data, we used PCA scores of 404 accessions to examine our phenotypic data structure. Confidence ellipses based on geographic areas (**Fig. 4a**; confidence level set to 95% around area centroid) indicated a degree of separation along the first PC for all four geographic areas based on combined phenotypes, most clearly separating 'India' and 'East Africa'. Accessions from 'Nepal' and 'Southern Africa' were relatively less differentiated compared to using genome-wide SNP data (see, e.g., F_{ST} values, **Tab. S5**). Trait-specific boxplots of phenotypic variation by region or genetic cluster (**Fig. 4b** and **Fig. S11**) illustrated a divergent pattern for some traits. For example, for plant height 'India' and 'East Africa' (or their corresponding clusters) showed the most difference, while for grain yield 'Nepal' and 'Southern Africa' did. On the other hand, for days to flowering, the greatest difference was between 'Nepal' and 'East Africa'.

Of the three last-mentioned traits, only plant height showed non-overlap between areas/clusters (for 'East Africa' vs. some other areas/clusters, where 'East Africa' accessions were, for example, on average 32% taller than accessions from 'India', **Fig. 4b**; the same non-overlap applied for the traits of leaf length, leaf number, leaf width [all greatest for 'East Africa'] and number of productive tillers [fewest for 'East Africa'], **Fig. S11**). In general, phenotypes were more differentiated between regions (Africa vs. Asia) than among geographic areas within regions, corresponding to genomic differentiation (**Fig. 2**) and consistent with combined phenotype centroids in **Fig. 4a**. Overall, each geographic area contained extensive phenotypic variation, with India containing the largest (**Tab. S7**). This is also be seen in the PCA and the boxplots (**Figs. 4** and **S11**).

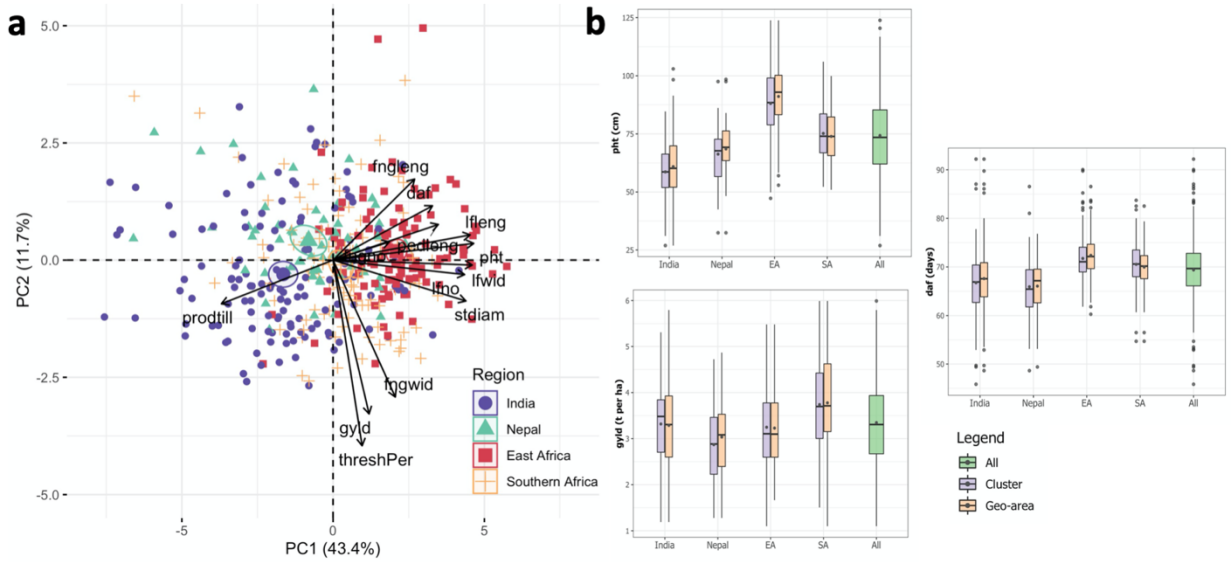


Fig. 4: Summary of phenotypic data for 423 finger millet accessions. a, The Principal component analysis (PCA) biplot shows vectors of each of the 13 phenotypic traits/variables (black arrows) and the PCA scores for combined phenotypes of 404 accessions (represented by points) coloured according to geographic areas; scores of accessions (19) without known location were excluded. The magnitude of the vectors shows the strength of their contribution to each PC. Vectors pointing in similar directions indicate positively correlated traits, vectors pointing in opposite directions indicate negatively correlated variables, and vectors at approximately right angles indicate low or no correlation. Coloured ellipses represent 95% confidence intervals around the centroid (bold symbol) for each area. stdiam, stem diameter; lfleng, leaf length; lfwid, leaf width; lfno, leaf number; fngleng, finger length; fngwid, finger width; fngno, finger number; pedleng, peduncle length; pht, plant height; prodtill, production tillers; daf, days to flowering; thershPer, threshing percentage; gyld, grain yield. **b**, Boxplots of the statistical distribution of individual trait values divided according to genetic clusters identified by Discriminant Analysis of Principal Components (purple boxplots), geographic areas (orange boxplots) and the total phenotypic variation (green boxplots). The middle bar and the point inside each boxplot represent median and mean, respectively. Examples are given for days to plant height (pht) and grain yield (gyld). Results for the other 10 phenotypic traits are shown in **Fig. S11**.

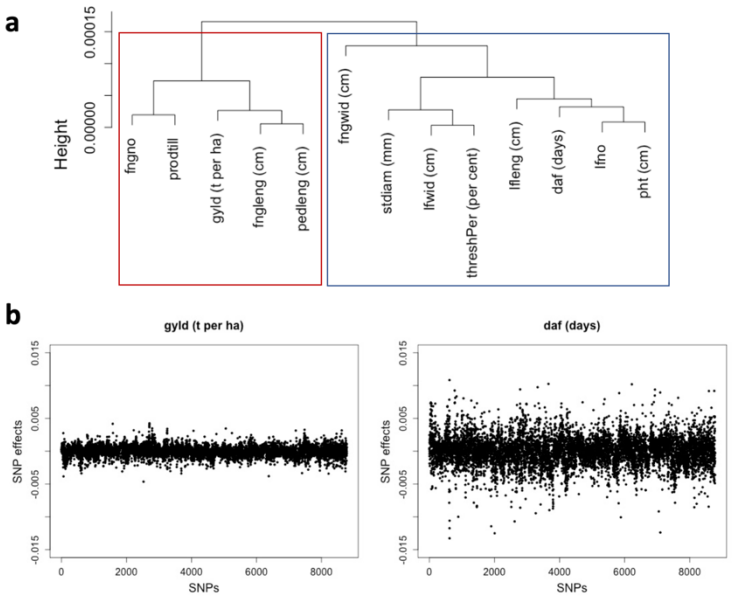
Genome-wide association analysis reveals a small number of significant associations and a range of candidate genes

Our GWAS detected 16 MTAs above the stringent Bonferroni threshold ($-\log_{10}(0.05/8{,}778)$ = 5.24), 15 of which were chromosome-located. Twelve were associated with finger length (seven on chromosome 2B, three on 5B, one on 7B and one on 8B), two with days to flowering (one on 4B and one scaffold marker) and two with threshing percentage (one each on 4B and 6B).

Manhattan and Q-Q plots for all 13 phenotypic traits are presented in **Fig. S12** and a list of all significant MTAs is given in **Tab. S8**. Most of the SNPs with significant MTAs that we identified by GWAS had a low MAF, in correspondence with our overall SNP panel (MAF < 0.2 for > 75% of all SNPs; see **Fig. S1d**). Therefore, care should be taken when interpreting our findings. The relatively small number of MTAs we detected could reflect the overall limited number of SNPs generated with the current genotyping strategy. All of the significant MTAs that we did detect were associated with the B sub-genome, even though it had a lower SNP density than the A sub-genome (but the B sub-genome does have slower LD decay, see above). A list of 61 unique putative candidate genes revealed within a 25 kb interval both upstream and downstream of significant MTAs is given in **Tab. S8**, but we do not here explore these associations further.

SNP effects clustering demonstrates different genetic architectures of phenotypic traits

Our exploration of SNP effects for phenotypic traits from RR-BLUP revealed two clusters of traits of different levels of complexity (**Fig. 5** and **Fig. S13**). Among traits with simpler genetic architecture were days to flowering and threshing percentage (consistent with the small number of MTAs detected for these two traits, see above). These are traits that do not follow a uniform distribution. Included among the traits with more complex genetic architecture (and of more uniform distribution) were finger length (which did still reveal MTAs in GWAS analysis) and grain yield (no MTAs in our analysis and known in other cereals to be highly polygenic; e.g., for wheat, see Brinton & Uauy, 2019).

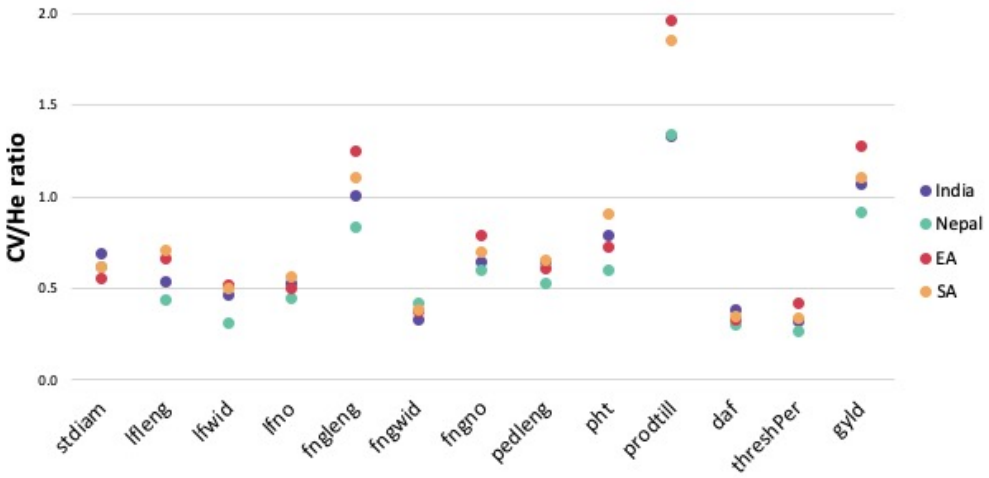


485 **Fig. 5: Clustering of SNP effects obtained from RR-BLUP to determine the complexity of**
486 **agronomic traits. a**, The UPGMA clustering algorithm was applied to the five statistical moments
487 of SNP effects for all 13 agronomic traits. The cluster in red consists of traits with highly polygenic
488 genetic architecture, and the cluster in blue consists of traits with simple- to moderately-complex
489 genetic architecture. **b**, Examples of SNP effect distributions for the complex trait of gyld (grain
490 yield) and the less complex trait daf (days to flowering). The results for the remaining 11 of the
491 individual traits evaluated are shown in **Fig. S13**.
492

493 **Comparison of phenotypic and genome-wide gene diversity by geographic area supports**
494 **varied post-domestication diversification pathways for finger millet**

495 Our geographic-area-based comparison of phenotypic trait diversity with overall underlying gene
496 diversity was based on the calculation of CV/H (i.e., standardized phenotypic variation per unit of
497 genome-wide gene diversity; **Fig. 6**). The results showed that the ranking of values between
498 geographic areas varied by trait, suggesting complex, context-specific diversification of the crop.
499 The comparison of rankings of CV/H for finger number and plant height, for example, showed
500 that the highest rank for the former was for 'East Africa' and for the latter was 'Southern Africa'.
501 Considering all 13 phenotypic traits, 'East Africa' most often of any geographic area ranked top
502 for CV/H (in 6 cases) and 'Nepal' bottom (11 cases). This is consistent with the relatively low

503 denominator (H) value for 'East Africa' compared to 'Nepal' (see above), and possibly indicates
504 an 'overexpression' of phenotypic variation in the former region that is consistent with a longer
505 domestication history. Days to flowering was the trait with the least spread in CV/H values for
506 geographic areas, indicating that variation in this trait may be the best phenotypic proxy of
507 underlying genomic diversity within geographic areas.



508 **Fig. 6 The ratio of the coefficient of phenotypic variance (CV) and genome-wide gene**
509 **diversity (H) for 13 agronomic traits for four geographic areas.** We used the ratios to inform
510 us of finger millet's selection history.

511

512 **DISCUSSION**

513 Broadening crop production requires more focus on orphan crops. Orphan crops have an intrinsic
514 value, but they can also provide broader lessons on crop domestication and crop diversification
515 pathways (Dawson et al., 2019). Here, we have studied the orphan crop finger millet and have
516 shown that a combination of modern and traditional methods can provide important insights
517 relevant for the future development of the crop. Our approach also provides a model for work on
518 other orphan crop species.

519 Past DNA-based studies of finger millet genetic diversity have generally been limited in scope,
520 involving < 150 accessions and/or < 100 polymorphic loci. These studies include those of Dida et
521 al. (2008), Arya et al. (2013), Manyasa et al. (2015), Ramakrishnan et al. (2016), Babu et al.
522 (2018), Lule et al. (2018) and Pandian et al. (2018) who all applied simple sequence repeat (SSR)
523 markers; and those of Gimode et al. (2016) and Sood et al. (2016) who used SSRs and SNPs.
524 Despite their limited scope, these previous studies revealed a degree of locally- and regionally-
525 structured genetic variation, including between Africa and Asia, and admixture between the two
526 continents.

527 A restricted number of previous studies have also sought to explore marker-trait associations in
528 finger millet. However, again the accessions and/or molecular markers involved were generally
529 limited to small numbers, and studies did not have available a finger millet genome assembly to
530 map markers to chromosome positions. These studies include those of Dida et al. (2021) who used
531 a medium-sized panel of SNPs but only 52 East African accessions to search for blast disease
532 resistance MTAs; Puranik et al. (2020) who used a large panel of SNPs but assessed only 190
533 genotypes to search for grain nutrient-content-related MTAs; Sharma et al. (2018) who applied a
534 medium-sized panel of SNPs to only 113 accessions to search for MTAs in 14 agro-morphological
535 traits; and Tiwari et al. (2020) who, beginning with the same SNPs and germplasm panel of Sharma
536 et al. (2018), explored marker associations with grain protein content. Despite their limited scope,
537 these previous studies indicated some MTAs. These were identified under varying degrees of
538 stringency and certainty levels, applying for example candidate gene sequence homologies for
539 characterization.

540 Our analysis was based on both an extensive germplasm panel and a broad genome-wide (and sub-
541 genome-located) SNP marker set. This was combined with extensive field phenotypic assessment,

the use of a finger millet genome assembly, and the application of a suite of population genomic tools. Overall, this provides us with a more complete picture than has been possible previously of genetic variation in the finger millet crop. Our analysis supports previous studies that have identified geographically-structured variation. However, it also sheds more light on the apparently complex evolutionary history of finger millet and its varied within and between region diversification pathways, thereby providing new information to support future breeding. Below, we focus further discussion on two aspects of particular interest that expand our current understanding of finger millet: i) polyploidy and sub-genome-specific diversification; and ii) geographically-specific relationships between phenotype and genotype embracing chromosome-, sub-genome-, and genome-level diversity, with a range of individual phenotypic traits and combinations of traits.

Sub-genome-specific diversification and polyploidy

Hybridization between different genomes, followed by chromosome doubling to generate polyploids, has played a crucial role in cereal crop evolution (Levy & Feldman, 2002). This has been well studied for major cereals such as allohexaploid wheat (Feldman & Levy, 2012). However, the full origin of the finger millet allotetraploid genome is not well understood (Zhang et al., 2019). The allopolyploidization event from diploid A and B sub-genome progenitors to form the tetraploid crop ancestor (*E. corocana* subsp. *africana*) is believed to have occurred in East Africa (de Wet et al., 1984). Traditionally, such polyploidization events are thought to be rare in plant's histories and initially lead to genomic bottlenecks in the derived organisms (Stebbins, 1950). In the current study, we could not compare relative diversity levels of the cultivated finger millet with the descendants of its wild tetraploid progenitor or previous diploid progenitors. However, previous studies found genomic diversity in cultivated finger millet to be significantly

lower than in wild *E. coracana* subsp. *africana* (e.g., Gimode et al., 2016). We were nevertheless able to explore for the first time in finger millet the sub-genomic and chromosome-specific diversity patterns in the crop by geographic area of sampling, which may inform on its evolution. We detected the highest level of genetic diversity in the B sub-genome overall in 'Southern Africa', with exceptionally high diversity levels along two chromosomes (5B and 9B). In contrast, for the A sub-genome, 'Southern Africa' only ranked third in diversity, while diversity for 'East Africa' ranked lowest for both sub-genomes.

Our new observations are consistent with broadening of diversity in specific parts of the finger millet genome as the crop developed new adaptations during expansion from East Africa to new areas in Africa and in Asia. This may have been facilitated by the polyploidization process that enables mechanisms such as inter-genomic transfer through translocation, recombination, and transposition to support rapid evolution (as outlined by Levy & Feldman, 2002). Our results could also indicate a secondary contact in the 'Southern Africa' region between the crop and its immediate progenitor, or perhaps new introgressions from other co-located *Eleusine* species that have specifically targeted the B sub-genome and could have been aided by polyploidization. Both wild *E. coracana* subsp. *africana* (i.e., the considered most likely immediate progenitor) and a range of other *Eleusine* species are sympatric with the finger millet crop in southern Africa (**Fig. 1**), providing opportunities for introgression. This type of process has been observed in other cereals, including wheat (He et al., 2019), and would be consistent with genotyping-by-sequencing in finger millet that has revealed chromosomal rearrangements between sub-genomes (Qi et al., 2018). Regardless of the cause, the patterns of sub-genome-specific diversity observed in our study suggest complexity in finer millet's evolution. Genomic exploration of *Eleusine* germplasm panels

587 containing extant descendants of the known and putative crop wild progenitors may further
588 elucidate this complex demographic history.

589 **Region-specific genomic-phenotypic relationships**

590 Our analysis shows complex relationships between genomic and phenotypic variation in finger
591 millet across the four geographic areas we sampled. This is consistent with a crop that has been
592 subject to millennia of domestication and has experienced different selection pressures based on
593 particular human preferences and production environments in different locations. Complex
594 relationships between genetic variation and phenotypic variation are common in crops (e.g., Kozak
595 et al., 2011), and our results were therefore not unexpected.

596 Our analysis of genetic diversity along sub-genomic chromosome homeologs indicated differences
597 in relative diversity for particular geographic areas. The subsequent pooling of chromosome
598 diversity at the sub-genome level also revealed different diversity rankings by location. When
599 levels of geographic-area-based phenotypic variation were compared with genome-wide gene
600 diversity (through the calculation of CV/H), the results suggested context-, trait-, and trait-
601 combination-specific diversification pathways for the crop. In addition, the 'East Africa' area most
602 often ranked top for CV/H values, consistent with a longer cultivation history here than elsewhere.
603 The profiles of phenotypic traits alone that varied between geographic areas by individual trait
604 further suggest multiple selection pressures both between and within areas, with the broad
605 variation often captured within areas indicating local differential selection.

606 Finally, patterns of differentiation between geographic areas also varied for genomic data
607 compared to phenotypic data, further supporting complex demographic processes. For example,
608 genome-wide SNPs revealed the highest differentiation between 'Southern Africa' and 'Nepal'

609 accessions, but combined phenotypic data indicated the greatest level of difference between 'India'
610 and 'East Africa'. This perhaps illustrates that these last two areas are the strongest foci for context-
611 specific crop development and/or are most agro-ecologically differentiated.

612 **Concluding remarks**

613 This study has revealed distinct genomic and phenotypic variation patterns in finger millet that
614 suggest complex diversification pathways for the crop. Our data provide a firmer basis for the
615 future development of finger millet. The information can, for example, help guide the
616 determination of heterotic groups to produce potentially hybrid finger millet varieties in the future
617 (Labroo et al., 2021; Mackay et al., 2021). Hybridization between 'East Africa' and 'Nepal' or
618 'Southern Africa' accessions may also generate useful genetic variation for use in the genetically
619 narrow East Africa region. A study of *Eleusine* germplasm panels containing known and putative
620 crop wild progenitors is required to understand the crop's demographic history further. In our
621 ongoing research, we seek to understand the nature and extent of genotype-by-environment
622 interactions in finger millet, to guide whether more local germplasm panels would be better suited
623 for GWAS, and to assess the prospects for climate change adaption of the crop. As genomic
624 resources continue to develop for finger millet, breeding material selection approaches that assign
625 genomic estimated breeding values to each separate sub-genome also open a possibility for
626 weighted sub-genome selection (Santantonio et al., 2019).

627 **ACKNOWLEDGEMENTS**

628 The authors are thankful for helpful comments from Ian Mackay and Rajiv Sharma, and
629 technical assistance from Ann Murithi and Erick Owuor Mikwa. We also thank Peter Bradbury

for adding a bespoke filtering option in TASSEL for removing SNPs with a heterozygosity level greater than $2pq \times (1 - F)$.

AUTHOR CONTRIBUTIONS

JB, IKD and DAO conceptualized the study. HFO coordinated the collection of field phenotype data. DAO coordinated the generation and initial screening of molecular marker data. SMJ carried out genomic DNA extractions of the 458 samples. JB conducted further data screening and all statistical analyses. IKD, RCG, JBu, GG and DAO contributed to determining appropriate analysis methods and interpreting findings. JB and IKD wrote the first draft of the paper that was then contributed to by all other authors.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

SUPPLEMENTAL MATERIAL

Supplemental File S1 contains supplemental tables (**Tables S2,5,6**) and figures (**Figs. S1-13**). **Supplemental File S2** contains supplemental **Tables S1,3,4,7,8**.

FUNDING

SRUC authors are grateful for Global Challenge Research Funding on orphan crops (project BB/P022537/1: Formulating Value Chains for Orphan Crops in Africa, 2017-2019, Foundation Award for Global Agriculture and Food Systems). JB was funded through an SRUC studentship Research Excellence Grant.

DATA AVAILABILITY

All data generated and analyzed during this study are included in this published article and its supplemental material. The raw phenotypic and DArTseq SNP data were deposited into Dryad Digital Repository (http://datadryad.org/).

REFERENCES

Arya, L., Verma, M., Gupta, V. K., & Seetharam, A. (2013). Use of genomic and genic SSR markers for assessing genetic diversity and population structure in Indian and African finger millet (Eleusine coracana (L.) Gaertn.) germplasm. *Plant Systematics and Evolution*, *299*(7), 1395–1401. https://doi.org/10.1007/s00606-013-0822-x

Babu, B. K., Sood, S., Chandrashekara, C., Pattanayak, A., & Kant, L. (2018). Mapping quantitative trait loci for important agronomic traits in finger millet (Eleusine coracana) mini core collection with genomic and genic SSR markers. *Journal of Plant Biochemistry and Biotechnology*, *27*(4), 401–414. https://doi.org/10.1007/s13562-018-0449-7

Bančič, J., Werner, C., Gaynor, C., Gorjanc, G., Odeny, D., Ojulong, H. F., … Hickey, J. M. (2021). Modelling illustrates that genomic selection provides new opportunities for intercrop breeding. *Frontiers in Plant Science*, *12*, 1–16. https://doi.org/10.1101/2020.09.11.292912

Bharathi, A. (2011). *Phenotypic and genotypic diversity if global finger millet (Eleusine coracana (L.) Gaertn.) composite collection*. Tamil Nadu Agricultural University.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Brinton, J., & Uauy, C. (2019). A reductionist approach to dissecting grain weight and yield in wheat. *Journal of Integrative Plant Biology*, *61*(3), 337–358. https://doi.org/10.1111/jipb.12741

Brooker, R. W., Jones, H. G., Paterson, E., Watson, C., Brooker, R. W., Bennett, A. E., … Mckenzie, B. M. (2015). Improving intercropping : a synthesis of research in agronomy , plant physiology and Research review Improving intercropping : a synthesis of research in agronomy , plant physiology and ecology. *New Phytologist*, *206*(APRIL), 107–117.

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., & Thompson, R. (2017). ASReml-R Reference Manual Version 4. *ASReml-R Reference Manual*, 176. Retrieved from https://asreml.kb.vsni.co.uk/wp-content/uploads/sites/3/2018/02/ASReml-R-Reference-Manual-4.pdf

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R.

683 (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.
684 https://doi.org/10.1093/bioinformatics/btr330

685 Dawson, I. K., Powell, W., Hendre, P., Bančič, J., Hickey, J. M., Kindt, R., … Jamnadass, R.
686 (2019). The role of genetics in mainstreaming the production of new and orphan crops to
687 diversify food systems and support human nutrition. *New Phytologist*.
688 https://doi.org/10.1111/nph.15895

689 de Wet, J. M. J., Prasada Rao, K. E., Brink, D. E., & Mengesha, M. H. (1984). Systematics and
690 Evolution of Eleusine coracana (Gramineae). *American Journal of Botany*, *71*(4), 550–
691 557.

692 Dida, M. M., & Devos, K. M. (2006). Finger Millet. *Cereals and Millets*, *1*, 333–343.

693 Dida, M. M., Oduori, C. A., Manthi, S. J., Avosa, M. O., Mikwa, E. O., Ojulong, H. F., &
694 Odeny, D. A. (2021). Novel sources of resistance to blast disease in finger millet. *Crop*
695 *Science*, *61*(1), 250–262. https://doi.org/10.1002/csc2.20378

696 Dida, M. M., Wanyera, N., Harrison Dunn, M. L., Bennetzen, J. L., & Devos, K. M. (2008).
697 Population Structure and Diversity in Finger Millet (Eleusine coracana) Germplasm.
698 *Tropical Plant Biology*, *1*(2), 131–141. https://doi.org/10.1007/s12042-008-9012-3

699 Dray, S., & Dufour, A. B. (2007). The ade4 package: Implementing the duality diagram for
700 ecologists. *Journal of Statistical Software*, *22*(4), 1–20.
701 https://doi.org/10.18637/jss.v022.i04

702 Endelman, J. B., & Jannink, J. L. (2012). Shrinkage estimation of the realized relationship
703 matrix. *G3: Genes, Genomes, Genetics*, *2*(11), 1405–1413.
704 https://doi.org/10.1534/g3.112.004259

705 Feldman, M., & Levy, A. A. (2012). Genome evolution due to allopolyploidization in wheat.
706 *Genetics*, *192*(3), 763–774. https://doi.org/10.1534/genetics.112.146316

707 Flint-Garcia, S. A., Thornsberry, J. M., & Edward IV, S. B. (2003). Structure of Linkage
708 Disequilibrium in Plants. *Annual Review of Plant Biology*, *54*, 357–374.
709 https://doi.org/10.1146/annurev.arplant.54.031902.134907

710 Gimode, D., Odeny, D. A., De Villiers, E. P., Wanyonyi, S., Dida, M. M., Mneney, E. E., … De
711 Villiers, S. M. (2016). Identification of SNP and SSR markers in finger millet using next
712 generation sequencing technologies. *PLoS ONE*, *11*(7), 1–23.
713 https://doi.org/10.1371/journal.pone.0159437

714 Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics.
715 *Molecular Ecology Notes*, *5*, 184–186. https://doi.org/10.1111/j.1471-8278

716 Hatakeyama, M., Aluri, S., Balachadran, M. T., Sivarajan, S. R., Patrignani, A., Grüter, S., …
717 Shimizu, K. K. (2018). Multiple hybrid de novo genome assembly of finger millet, an
718 orphan allotetraploid crop. *DNA Research*, *25*(1), 39–47.
719 https://doi.org/10.1093/dnares/dsx036

He, F., Pasam, R., Shi, F., Kant, S., Keeble-Gagnere, G., Kay, P., … Akhunov, E. (2019). Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. *Nature Genetics*, *51*(5), 896–904. https://doi.org/10.1038/s41588-019-0382-2

Hill, W. G., & Weir, B. S. (1988). Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology*, *33*(1), 54–78. https://doi.org/10.1016/0040-5809(88)90004-4

Hilu, K. W., & de Wet, J. M. J. (1976). Domestication of Eleusine coracana. *Economic Botany*, *30*(3), 199–208. Retrieved from https://www.jstor.org/stable/4253732

Hittalmani, S., Mahesh, H. B., Shirke, M. D., Biradar, H., Uday, G., Aruna, Y. R., … Mohanrao, A. (2017). Genome and Transcriptome sequence of Finger millet (Eleusine coracana (L.) Gaertn.) provides insights into drought tolerance and nutraceutical properties. *BMC Genomics*, *18*(1), 1–16. https://doi.org/10.1186/s12864-017-3850-z

Jaiswal, V., Gupta, S., Gahlaut, V., Muthamilarasan, M., Bandyopadhyay, T., Ramchiary, N., & Prasad, M. (2019). Genome-Wide Association Study of Major Agronomic Traits in Foxtail Millet (Setaria italica L.) Using ddRAD Sequencing. *Scientific Reports*, *9*(1), 3–5. https://doi.org/10.1038/s41598-019-41602-6

Jamnadass, R., Mumm, R. H., Hale, I., Hendre, P., Muchugi, A., Dawson, I. K., … Van Deynze, A. (2020). Enhancing African orphan crops with genomics. *Nature Genetics*, *52*(4), 356–360. https://doi.org/10.1038/s41588-020-0601-x

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kamenya, S. N., Mikwa, E. O., Song, B., & Odeny, D. A. (2021). Genetics and breeding for climate change in Orphan crops. *Theoretical and Applied Genetics*, 1–29. https://doi.org/10.1007/s00122-020-03755-1

Kassambara, A., & Mundt, F. (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. Retrieved from https://cran.r-project.org/package=factoextra

Komsta, L., & Novomestky, F. (2011). Package ‘ moments ,’ 1–15.

Kooke, R., Kruijer, W., Bours, R., Becker, F., Kuhn, A., van de Geest, H., … Keurentjes, J. J. B. (2016). Genome-wide association mapping and genomic prediction elucidate the genetic architecture of morphological traits in arabidopsis. *Plant Physiology*, *170*(4), 2187–2203. https://doi.org/10.1104/pp.15.00997

Kozak, M., Bocianowski, J., Liersch, A., Tartanus, M., Bartkowiak-Broda, I., Piotto, F. A., & Azevedo, R. A. (2011). Genetic divergence is not the same as phenotypic divergence. *Molecular Breeding*, *28*(2), 277–280. https://doi.org/10.1007/s11032-011-9583-9

Labroo, M. R., Studer, A. J., & Rutkoski, J. E. (2021). Heterosis and Hybrid Crop Breeding : A Multidisciplinary Review, *12*(February), 1–19. https://doi.org/10.3389/fgene.2021.643761

Levy, A. A., & Feldman, M. (2002). The impact of polyploidy on grass genome evolution. *Plant Physiology*, *130*(4), 1587–1593. https://doi.org/10.1104/pp.015727

Liu, Q., Jiang, B., Wen, J., & Peterson, P. M. (2014). Low-copy nuclear gene and McGISH resolves polyploid history of eleusine coracana and morphological character evolution in Eleusine. *Turkish Journal of Botany*, *38*(1), 1–12. https://doi.org/10.3906/bot-1305-12

Lule, D., De Villiers, S., Fetene, M., Odeny, D. A., Rathore, A., Das, R. R., & Tesfaye, K. (2018). Genetic diversity and association mapping of Ethiopian and exotic finger millet accessions. *Crop and Pasture Science*, *69*(9), 879–891. https://doi.org/10.1071/CP18175

Mackay, I. J., Cockram, J., Howell, P., & Powell, W. (2021). Understanding the classics: the unifying concepts of transgressive segregation, inbreeding depression and heterosis and their central relevance for crop breeding. *Plant Biotechnology Journal*, *19*(1), 26–34. https://doi.org/10.1111/pbi.13481

Mangin, B., Siberchicot, A., Nicolas, S., Doligez, A., This, P., & Cierco-Ayrolles, C. (2012). Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity*, *108*(3), 285–291. https://doi.org/10.1038/hdy.2011.73

Manyasa, E. O., Tongoona, P., Shanahan, P., Mgonja, M. A., & De Villiers, S. (2015). Genetic diversity in East African finger millet (Eleusine coracana (L.) Gaertn) landraces based on SSR markers and some qualitative traits. *Plant Genetic Resources: Characterization and Utilization*, *13*(1), 45–55. https://doi.org/10.1017/S1479262114000628

Manyasa Okuku, E., Tongoona, P., Shanahan, P., Githiri, M., & Rathore, A. (2016). Correlations, path coefficient analysis and heritability for quantitative traits in finger millet landraces. *Philippine Journal of Science*, *145*(2), 197–208.

Marroni, F., Pinosio, S., Zaina, G., Fogolari, F., Felice, N., Cattonaro, F., & Morgante, M. (2011). Nucleotide diversity and linkage disequilibrium in Populus nigra cinnamyl alcohol dehydrogenase (CAD4) gene. *Tree Genetics and Genomes*, *7*(5), 1011–1023. https://doi.org/10.1007/s11295-011-0391-5

Meyer, R. S., & Purugganan, M. D. (2013). Evolution of crop species: Genetics of domestication and diversification. *Nature Reviews Genetics*, *14*(12), 840–852. https://doi.org/10.1038/nrg3605

Mirza, N., & Marla, S. S. (2019). Finger Millet (Eleusine coracana L. Gartn.) Breeding. In J. M. Al-Khayri, S. M. Jain, & D. V Johnson (Eds.), *Advances in Plant Breeding Strategies: Cereals: Volume 5* (pp. 83–132). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-23108-8_3

Mustafa, M. A., Mayes, S., & Massawe, F. (2019). Crop Diversification Through a Wider Use of Underutilized Crops: A Strategy to Ensure Food and Nutrition Security in the Face of Climate Change. In A. Sarkar, S. R. Sensarma, & G. W. vanLoon (Eds.), *Sustainable Solutions for Food Security : Combating Climate Change by Adaptation* (pp. 125–149). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-77878-5_7

795 Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of
796 restriction endonucleases. *Proceedings of the National Academy of Sciences of the United*
797 *States of America*, *76*(10), 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

798 Nei, Masatoshi. (1973). Analysis of Gene Diversity in Subdivided Populations. *Proceedings of*
799 *the National Academy of Sciences of the United States of America*, *70*(12), 3321–3323.
800 https://doi.org/10.1073/pnas.70.12.3321

801 Odeny, D. A., Niazi, A., Tesfaye, K., Lule, D., Wanyonyi, S., & Kunguni, J. S. (2020). Genomic
802 Designing for Climate Smart Finger Millet. In C. Kole (Ed.), *Genomic Designing of*
803 *Climate-Smart Cereal Crops* (pp. 287–307). Cham: Springer International Publishing.
804 https://doi.org/10.1007/978-3-319-93381-8_7

805 Pandian, S., Satish, L., Rameshkumar, R., Muthuramalingam, P., Rency, A. S., Rathinapriya, P.,
806 & Ramesh, M. (2018). Analysis of population structure and genetic diversity in an exotic
807 germplasm collection of Eleusine coracana (L.) Gaertn. using genic-SSR markers. *Gene*,
808 *653*(February), 80–90. https://doi.org/10.1016/j.gene.2018.02.018

809 Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular
810 approach. *Bioinformatics*, *26*(3), 419–420. https://doi.org/10.1093/bioinformatics/btp696

811 Pearson, K. (1895). VII. Note on regression and inheritance in the case of two parents.
812 *Proceedings of the Royal Society*, *58*, 240–242. Retrieved from
813 http://doi.org/10.1098/rspl.1895.0041

814 Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from
815 Genome-Wide Allele Frequency Data. *PLoS Genetics*, *8*(11).
816 https://doi.org/10.1371/journal.pgen.1002967

817 Puranik, S., Sahu, P. P., Beynon, S., & Srivastava, R. K. (2020). Genome-wide association
818 mapping and comparative genomics identifies genomic regions governing grain nutritional
819 traits in finger millet (Eleusine coracana L . Gaertn .). *Plants People Planet*, (July), 1–14.
820 https://doi.org/10.1002/ppp3.10120

821 Qi, P., Gimode, D., Saha, D., Schröder, S., Chakraborty, D., Wang, X., … Devos, K. M. (2018).
822 UGbS-Flex, a novel bioinformatics pipeline for imputation-free SNP discovery in
823 polyploids without a reference genome: Finger millet as a case study. *BMC Plant Biology*,
824 *18*(1), 1–19. https://doi.org/10.1186/s12870-018-1316-3

825 R Core Team. (2019). R: A Language and Environment for Statistical Computing. Vienna,
826 Austria. Retrieved from https://www.r-project.org/

827 Ramakrishnan, M., Antony Ceasar, S., Duraipandiyan, V., Al-Dhabi, N. A., & Ignacimuthu, S.
828 (2016). Assessment of genetic diversity, population structure and relationships in Indian and
829 non-Indian genotypes of finger millet (Eleusine coracana (L.) Gaertn) using genomic SSR
830 markers. *SpringerPlus*, *5*(1), 1–11. https://doi.org/10.1186/s40064-015-1626-y

831 Rao, S. A. (1980). *Germplasm collecting mission to Zambia*. Patancheru, Andhra Pradesh, India.
832 Retrieved from http://eprints.icrisat.ac.in/id/eprint/4229

833 Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., &
834 Mesirov, J. P. (2011). Integrative Genome Viewer. *Nature Biotechnology*, *29*(1), 24–26.
835 https://doi.org/10.1038/nbt.1754.Integrative

836 Rodríguez, G. R., Muños, S., Anderson, C., Sim, S. C., Michel, A., Causse, M., … van der
837 Knaap, E. (2011). Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm
838 and the relationship to fruit shape diversity. *Plant Physiology*, *156*(1), 275–285.
839 https://doi.org/10.1104/pp.110.167577

840 Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing
841 phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425.
842 https://doi.org/10.1093/oxfordjournals.molbev.a040454

843 Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., & Kilian, A.
844 (2011). Diversity Arrays Technology (DArT) and next-generation sequencing combined:
845 genome-wide, high throughput, highly informative genotyping for molecular breeding of
846 Eucalyptus. *BMC Proceedings*, *5*(S7). https://doi.org/10.1186/1753-6561-5-s7-p54

847 Santantonio, N., Jannink, J. L., & Sorrells, M. (2019). Prediction of subgenome additive and
848 interaction effects in allohexaploid wheat. *G3: Genes, Genomes, Genetics*, *9*(3), 685–698.
849 https://doi.org/10.1534/g3.118.200613

850 Schliep, K. P. (2011). phangorn: Phylogenetic analysis in R. *Bioinformatics*, *27*(4), 592–593.
851 https://doi.org/10.1093/bioinformatics/btq706

852 Sharma, D., Tiwari, A., Sood, S., Jamra, G., Singh, N. K., Meher, P. K., & Kumar, A. (2018).
853 Genome wide association mapping of agro-morphological traits among a diverse collection
854 of finger millet (Eleusine coracana L.) genotypes using SNP markers. *PLoS ONE*, *13*(8), 1–
855 21. https://doi.org/10.1371/journal.pone.0199444

856 Sood, S., Joshi, D. C., Chandra, A. K., & Kumar, A. (2019). Phenomics and genomics of finger
857 millet: current status and future prospects. *Planta*, *250*(3), 731–751.
858 https://doi.org/10.1007/s00425-019-03159-6

859 Sood, S., Kumar, A., Kalyana Babu, B., Gaur, V. S., Pandey, D., Kant, L., & Pattnayak, A.
860 (2016). Gene discovery and advances in finger millet [Eleusine coracana (L.) Gaertn.]
861 genomics-an important nutri-cereal of future. *Frontiers in Plant Science*, *7*, 1–17.
862 https://doi.org/10.3389/fpls.2016.01634

863 South, A. (2011). rworldmap: A new R package for mapping global data. *R Journal*, *3*(1), 35–43.
864 https://doi.org/10.32614/rj-2011-006

865 Stebbins, G. L. (1950). Variation and Evolution in Plants (p. 644). Boston, MA: Columbia
866 University Press. https://doi.org/https://doi.org/10.7312/steb94536

867 Tiwari, A., Sharma, D., Sood, S., Jaiswal, J. P., Pachauri, S. P., Ramteke, P. W., & Kumar, A.
868 (2020). Genome-wide association mapping for seed protein content in finger millet
869 (Eleusine coracana) global collection through genotyping by sequencing. *Journal of Cereal*
870 *Science*, *91*(May 2019), 102888. https://doi.org/10.1016/j.jcs.2019.102888

Upadhyaya, H. D., Gowda, C. L. L., Pundir, R. P. S., Reddy, V. G., & Singh, S. (2006). Development of core subset of finger millet germplasm using geographical origin and data on 14 quantitative traits. *Genetic Resources and Crop Evolution*, *53*(4), 679–685. https://doi.org/10.1007/s10722-004-3228-3

von Grebmer, K., Saltzman, A., Birol, E., Wiesmann, D., Prasai, N., Yin, S., … Menon, P. (2014). *2014 Global Hunger Index: The Challenge of Hidden Hunger*. Bonn, Washington, D.C., and Dublin: Welthungerhilfe, International Food Policy Research Institute, and Concern Worldwide. Retrieved from http://dx.doi.org/10.2499/9780896299580

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York. Retrieved from http://had.co.nz/ggplot2/book

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., … Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, *38*(2), 203–208. https://doi.org/10.1038/ng1702

Zhang, H., Hall, N., Goertzen, L. R., Chen, C. Y., Peatman, E., Patel, J., & Scott McElroy, J. (2019). Transcriptome analysis reveals unique relationships among eleusine species and heritage of eleusine coracana. *G3: Genes, Genomes, Genetics*, *9*(6), 2029–2036. https://doi.org/10.1534/g3.119.400214