RESEARCH ARTICLE

Genetic diversity, population structure and selection signatures in Enset (*Ensete ventricosum*, (Welw.) Cheesman), an underutilized and key food security crop in Ethiopia

**Alye Tefera Haile · Mallikarjuna Rao Kovi ·
Sylvia Sagen Johnsen · Bizuayehu Tesfaye ·
Trine Hvoslef-Eide · Odd Arne Rognli**

Abstract Enset (*Ensete ventricosum* (Welw.) Cheesman) is a multipurpose, drought-tolerant and a key food security crop, which is the staple food for peoples in the south and southwestern parts of Ethiopia. In this study, we characterized genetic diversity, population structure and selection signatures in 226 cultivated and 10 wild enset accessions collected from diverse enset growing regions of Ethiopia using 3505 high-quality single nucleotide polymorphism (SNP) markers obtained from double-digest restriction-site associated DNA (ddRAD) sequences. The population structure and cluster analyses clearly distinguished between cultivated and wild enset. The analysis of molecular variance (AMOVA) revealed higher levels of genetic diversity within populations and regions (91.2 and 92.4%, respectively) than between populations and regions (8.8 and 7.6%, respectively). This shows that the region of origin and environmental heterogeneity have little influence on the genetic diversity. However, the genetic differentiation between regions was moderate to large ($F_{ST} = 0.06–0.17$). The genetic structure of enset was mainly shaped by eco-geographic factors, mode of propagation and cultivation status. Six genes potentially involved in sexual reproduction and flowering signalling, which are key processes underlying domestication and adaptation, were under positive selection demonstrating that sexual reproduction plays an important role in shaping enset diversity. A lot of unexplored diversity is available for improving enset in Ethiopia, with patterns of diversity consistent with divergent selection on adaptive traits. This diversity also shows potential for introducing enset as a more food secure crop for the food insecure regions in the dry north of Ethiopia.

A. T. Haile · M. R. Kovi · S. S. Johnsen · T. Hvoslef-Eide ·
O. A. Rognli (✉)
Department of Plant Sciences, Faculty of Biosciences,
Norwegian University of Life Sciences, P.O. Box 5003,
1432 Ås, Norway
e-mail: odd-arne.rognli@nmbu.no

A. T. Haile · B. Tesfaye
School of Plant and Horticultural Science, Faculty
of Agriculture, Hawassa University, P.O. Box 05, Awassa,
Ethiopia

Introduction

Enset (*Ensete ventricosum*, (Welw.) Cheesman), commonly known as false banana or Abyssinian banana, is a perennial diploid ($2n = 18$), monocarpic species belonging to the family *Musaceae* in the genus *Ensete* (Westphal et al. 1975). Enset, banana and plantains are the most important cultivated members of the family; all with high global and local economic as

well as food security importance (Baker and Simmonds 1953; Simmonds 1962). The genus *Ensete* consists of three species with extensive geographical distribution, *E. ventricosum* and *E. livingstonianum* in Africa and *E. glaucum* in Asia, and five other localized endemics or near-endemic species (Borrell et al. 2019). Enset (*Ensete ventricosum*, (Welw.) Cheesman) is the only cultivated species of the genus *Ensete* and its domestication and utilization as a food and fiber crop is so far restricted to Ethiopia. It was domesticated in Ethiopia as early as 10,000 years ago (Brandt et al. 1997). Enset is highly drought tolerant with a wide agroecological distribution and is cultivated only with household-produced inputs (Brandt et al. 1997; Tsegaye and Struik 2002). It is unknown whether its wide distribution across a range of altitudes involves genetic or phenotypic adaptation (Tsegaye 2002). Eighty percent of the enset production is concentrated in the southern and southwestern part of Ethiopia (Bezuneh et al. 1967), where it serves as a staple and co-staple food for about 25 mill people (Borrell et al. 2020; Brandt et al. 1997; Spring et al. 1996). Furthermore, it is used for several other purposes, such as animal feed, fiber, construction material and in traditional medicine. The crop grows best at cooler, higher altitudes and is found mostly between 1200–3100 m above sea level (Brandt et al. 1997). Enset plants grow 4–8 m, sometimes up to 11 m height. Cultivated enset are propagated vegetatively, while wild enset reproduces through seeds (Birmeta et al. 2004; Borrell et al. 2019; Tsegaye and Struik 2001). Enset is usually harvested 4–6 years after transplantation, but age at harvest varies between 3 to 12 years (Borrell et al. 2020; Brandt et al. 1997). Thus, if other crops fail, enset plants can be harvested at any time, providing security against hunger for farmers and their families. This became evident through the great famine in Ethiopia in the years 1888 to 1892 (Tobiaw and Bekele 2011), and is the reason why enset is called "The Tree Against Hunger" (Brandt et al. 1997; Costa 1984). This is an important aspect of introducing enset to other, more food insecure regions in Ethiopia, particularly in the dry north.

Ethiopia is the center of origin of many plant species, including enset (Engels and Hawkes 1991). The presence of wild and cultivated enset indicates that Ethiopia is the primary center of origin and center of diversity (Purseglove 1985; Vavilov 1951).

Ethnic groups in Ethiopia recognize and exploit various enset landraces. Regions in Ethiopia with diverse cultural history have rich biodiversity (Tsegaye 2002). Enset-based farming system is a major agricultural system and farmers cultivate many enset landraces across various climatic and agroecological systems (Borrell et al. 2019). Research on genetic diversity of specific enset accessions from local regions using molecular markers such as amplified fragment length polymorphism (AFLP) (Negash et al. 2002; Tesfamicael et al. 2020), random amplified polymorphic DNA (RAPD) (Birmeta et al. 2004), Inter simple sequence repeats (ISSR) (Tobiaw and Bekele 2011), chloroplast DNA sequences (Bekele and Shigeta 2011), simple sequence repeats (SSR) (Gerura et al. 2019; Getachew et al. 2014; Olango et al. 2015; Biswas et al. 2020; Nuraga et al. 2022) and single nucleotide polymorphisms (SNPs) (Tesfamicael et al. 2020) revealed genetic diversity among and within wild and cultivated enset accessions. SNP markers are powerful tools for estimating genetic similarities and diversity. SNP markers are abundant and robust, suitable for automated high-throughput genotyping of many samples and are able to resolve differences among extremely similar individuals and increase the accuracy of diversity estimates (Hinze et al. 2017). The double-digest restriction-site associated DNA (ddRAD) technique is a powerful and relatively cost-effective approach for developing numerous SNP markers and constructing high-density genetic maps (Peterson et al. 2012). It has been used extensively for population genetic research in a wide range of non-model organisms (Andrews et al. 2016; Peterson et al. 2012).

The cultivated enset is vegetatively propagated, genetic divergence among clones may be minimal and could be difficult to detect using these marker types (McKey et al. 2010). Moreover, different molecular markers have different properties and will reveal different aspects of genetic diversity (Karp et al. 1997). The investigations mentioned above were conducted in certain enset growing areas in the southern and southwestern part of the country. Since Ethiopia is the center of diversity, many enset rich locations harboring large amounts of diversity of cultivated and wild enset are yet to be studied and is not represented in ex situ collections. Enset clones have traditionally been characterized phenotypically, however, phenotypic description is limited by the cost, time and

space required to make visual observations and measurements (Hinze et al. 2017).

Despite the abundance, diversity and ecological importance of enset, the species is not well characterized at the genomic level and has been far less studied than other cultivated species in the family *Musaceae* (Borrell et al. 2019). More detailed diversity research of both cultivated and wild enset accessions in Ethiopia is necessary to meet future needs, including diversification of crops in more vulnerable regions in Ethiopia. Novel sources of genetic diversity need to be identified, characterized, incorporated into breeding programs, and utilized for the development of non-redundant core collections for conservation and breeding. In this study, SNP markers were developed and used to understand the population divergence of cultivated and wild enset. Understanding the genetic basis of enset domestication provides a valuable foundation for enset conservation and genetic improvement. The objectives of the present study were: (1) to evaluate the efficacy and suitability of SNP markers developed from ddRAD sequencing for high-throughput genotyping of enset; (2) to assess population structure, genetic diversity, and relationships among and within cultivated and wild enset accessions, and (3) to identify candidate genes potentially subjected to domestication and selection.

Materials and methods

Sampling area

The Southern Nations, Nationalities and Peoples' Region (SNNPR) state has a total area of 117,506 km^2, with altitudes ranging from 378 to 4,201 m above sea level (m a.s.l) (Abebe 2005). Enset accessions were collected from three main enset culture communities, which are densely populated enset cultivating administrative regions (Sidama, Gurage and South Omo). The wild enset were collected around farms, along riversides and in deep forests. The three collection regions were deliberately chosen based on their enset production potential in SNNPR, where more than two-thirds of the country's enset production is located (Zeberga et al. 2014). We collected 226 cultivated and 10 wild enset accessions originating from different geographical locations and agroecological zones (Table 1; Supplementary Table 1). The

Table 1 Enset (*Ensete. ventricosum* (Welw.) Cheesman) plant materials from Ethiopia used for genetic diversity analysis

Cultivation status	Number of accessions	Geographical origin and location	Sources
Cultivated	72	Sidama	Farmers
Cultivated	62	Gurage	Farmers
Cultivated	92	South Omo	Farmers
Wild	10	South Omo	Farmers and forest area

major ethnic regions cultivating enset and the study areas in Southern Ethiopia are show in Fig. 1.

Preparation of NaCl-CTAB preservation and samples collection

The saturated NaCl-CTAB solution was used to preserve the enset leaf samples upon collection, as described by Rogstad et al. (Rogstad 1992) with minor modifications. Briefly, 550 g NaCl was added to 1 L of water, boiled, and cooled at ambient temperature, and mixed thoroughly until the salt precipitated. Then, 35 g of CTAB was added gradually with intermittent irregular intervals mixing, until the solution became viscous. 35–40 mL of the prepared solution was aliquoted into 50 mL Falcon tubes and used for preservation of tissue samples. A pair of scissors was used to remove leaf samples from the mother plants, and the scissors were cleaned with ethanol (96%) between independent samples. Fresh cigar-leaf samples harvested from each enset accession were stored immediately in the 50 mL tubes containing the saturated NaCl-CTAB preservation solution. Samples were then placed in a black plastic bag and stored in a dark room at ambient temperature to preserve genomic DNA from degradation during transportation from the farmer fields in Ethiopia to the laboratory in Norway.

DNA extraction

DNA was extracted from the preserved leaf samples using the DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany). DNA quality and quantity were determined using a NanoDrop spectrophotometer
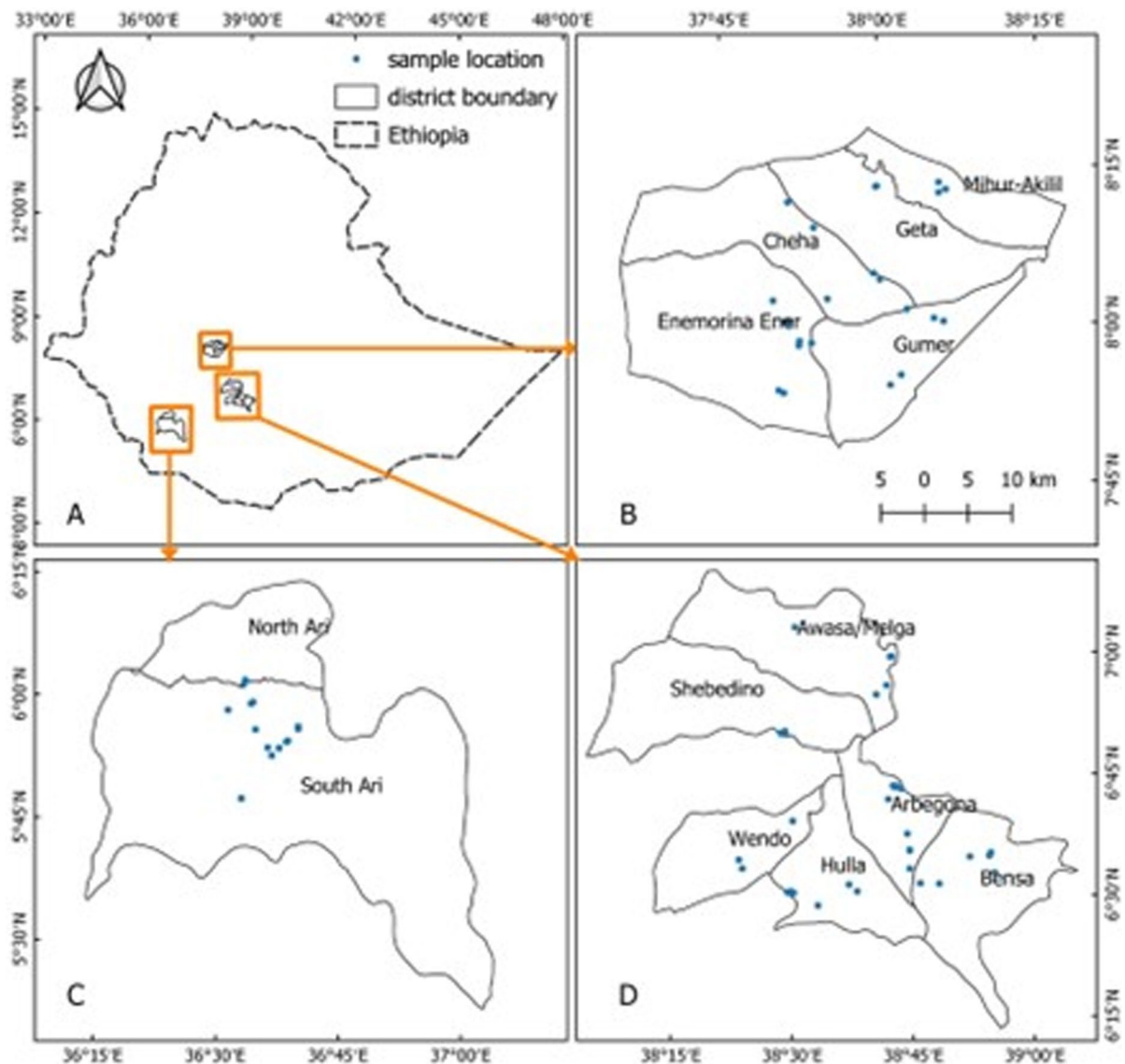
Fig. 1 An overview of the study districts in Ethiopia (**A**) and their detailed locations in Gurage (**B**), South Omo (**C**) and Sidama (**D**)

(Thermo Fisher, Inc.) and agarose gel electrophoresis (1%). DNA concentrations were determined using the Qubit® dsDNA BR assay kit (Life Sciences) and Quant-iT™ PicoGreen™ (Life Sciences) dsDNA assay.

Double-digest restriction-site-associated DNA (ddRAD) library preparation and Illumina sequencing

We calculated the number of reads required for 20X coverage of restriction fragments in the 150–500 bp

size range across 10 multiplexed individuals using multiple enzyme pairs, assuming 0.44 GC content, to ensure that restriction fragments could feasibly be sequenced with enough coverage on an Illumina MiSeq platform. The ddRAD procedure used in this study was modified from Peterson et al. (2012) (For further ddRAD information and the complete protocol, see Supplementary information, Table 2–7; Supplementary Table 1). 500 ng of each DNA sample was double digested using *EcoR1* HF (the "rare cutter"– recognizes a six bases motif, i.e., 5′-GAA

TTC), and *MseI* (the "frequent cutter"– recognizes a four bases motif, i.e., 5′-TTAA) restriction endonucleases, and adapters ligated to the digested fragments. Each DNA sample with a unique P1 barcode, and a P2 barcode common for all samples. Samples containing unique P1 barcodes were pooled, and the Sage Science Blue Pippin system (https://sagescience.com/) was used to select fragments of about 600 bp to reduce the possibility of unknown introns in the selected sequences and maximize the chances of obtaining SNPs. Size-selected libraries was bound to Dynabeads® M-270 Streptavidin magnetic beads (Invitrogen), to eliminate fragments without the P2 adapter, and the libraries amplified by PCR using Phusion™ Polymerase kit (Invitrogen) and index-marked primers for further tagging of the samples. The libraries were analyzed using an Agilent 2100 Bioanalyzer and diluted to a concentration of 35 nM for paired-end sequencing using the V2 sequencing kit on the MiSeq platform (Illumina). The sequencing was performed at the Norwegian University of Life Sciences, Norway.

Sequence data analysis and SNP calling

The GBS data obtained was quality checked using FASTQ format, FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). High quality reads were retained after trimming the bad quality reads using Trimmomatic program (Bolger et al. 2014). The raw paired-end sequence reads obtained from MiSeq were quality checked after removing the adapters and barcodes. The clean paired-end reads were used to call the SNPs using the STACKS 2 (Rochette et al. 2019) pipeline. The SNPs were filtered based on the following criteria: (1) variant should be bi-allelic SNPs, (2) SNPs having more than 20% missing information were excluded, (3) genotypes having more than 20% missing information were excluded, and (4) markers with minor allele frequency MAF > 0.05 were retained.

Population structure analysis

Population groups were inferred using the fastSTRU CTURE software (Raj et al. 2014). Twenty independent test runs were conducted allowing K to vary from 1 to 20. The optimal value of K for these runs was then determined using the ChooseK function. The script ChooseK, included with the fastSTRUCTURE package, was used to choose the number of subpopulations that maximize the marginal likelihood. The cluster membership matrices of the fastSTRUCTURE outputs were visualized using structure selector tool (Li and Liu 2018). Following the assignment of individuals to populations, the program package CLUMPACK (Kopelman et al. 2015) was used to summarize the structure results into structure plots.

Genetic diversity analysis

For genetic diversity analysis, subpopulations were defined as the number of clusters produced by fastSTRUCTURE at K = 12. Genetic diversity among and within populations, observed (H_O) and expected (H_E) heterozygosity, and pairwise fixation index (F_{ST}) for the subpopulations (Weir and Cockerham 1984) was estimated by analysis of molecular variance (AMOVA) using Arlequin v.3.5 (Excoffier and Lischer 2010). Significance (P < 0.05) of the F_{ST} values were estimated using 1023 permutations. F_{ST} results were interpreted using the same standard as in Pino Del Carpio et al. 2011; Hartl et al. 1997; and Wright 1978.

Phylogenetic trees and PCA analyses

To examine the relationship between cultivated and wild enset accessions, Principal Component Analysis (PCA) was performed using TASSEL v5.2 (Bradbury et al. 2007) and maximum-likelihood (ML) phylogenetic tree analyses performed using PhyML 3.0 (Guindon et al. 2010). The trees were prepared and visualized using the iTOL v4 online tool (Letunic and Bork 2019). PCAs were graphically summarized using scatter plots. Populations were named according to the passport data denoting geographical origin.

FST outlier tests for detecting SNP loci under selection

To detect loci under directional selection, we used the hierarchical method (Excoffier et al. 2009), a modified approach of Beaumont and Nichols (1996), implemented in the ARLEQUIN software package version 3.5.1.3 (Excoffier and Lischer 2010). We employed a hierarchical island model based on 2 groups (cultivated and wild enset) with 50,000

simulations to generate the joint distribution of F_{ST} versus heterozygosity. Loci that fall out of the 99% confidence intervals of the distribution were identified as outliers being putatively under selection. The putative function of genes with outlier SNPs was identified using the Gene Ontology (GO) annotation using the Blast2GO software tool version 3.0 (Conesa et al. 2005).

Results

SNP discovery and filtering

Following sequencing of the double digest RAD fragments, data processing and SNP filtering, the alleles with high heterozygosity (> 80%) were removed and a total number of 3505 high-quality SNPs were identified among the 236 enset accessions.

Fig. 2 Population structure (**A**) Model complexity that maximizes likelihood is 12 and the highest peak shows on the graph the best K = 12 (**B**) Population structure based on fastSTRUCTURE output resulting in K = 12 being the most likely number of genetic clusters, where each cluster is represented as a different shade and each bar represents an individual within each geographic region/cultivation status; colors represent the groups identified. 1: Gurage (cultivated enset); 2: Sidama (cultivated enset); 3: South Omo (cultivated enset); 4: wild enset (from South Omo)
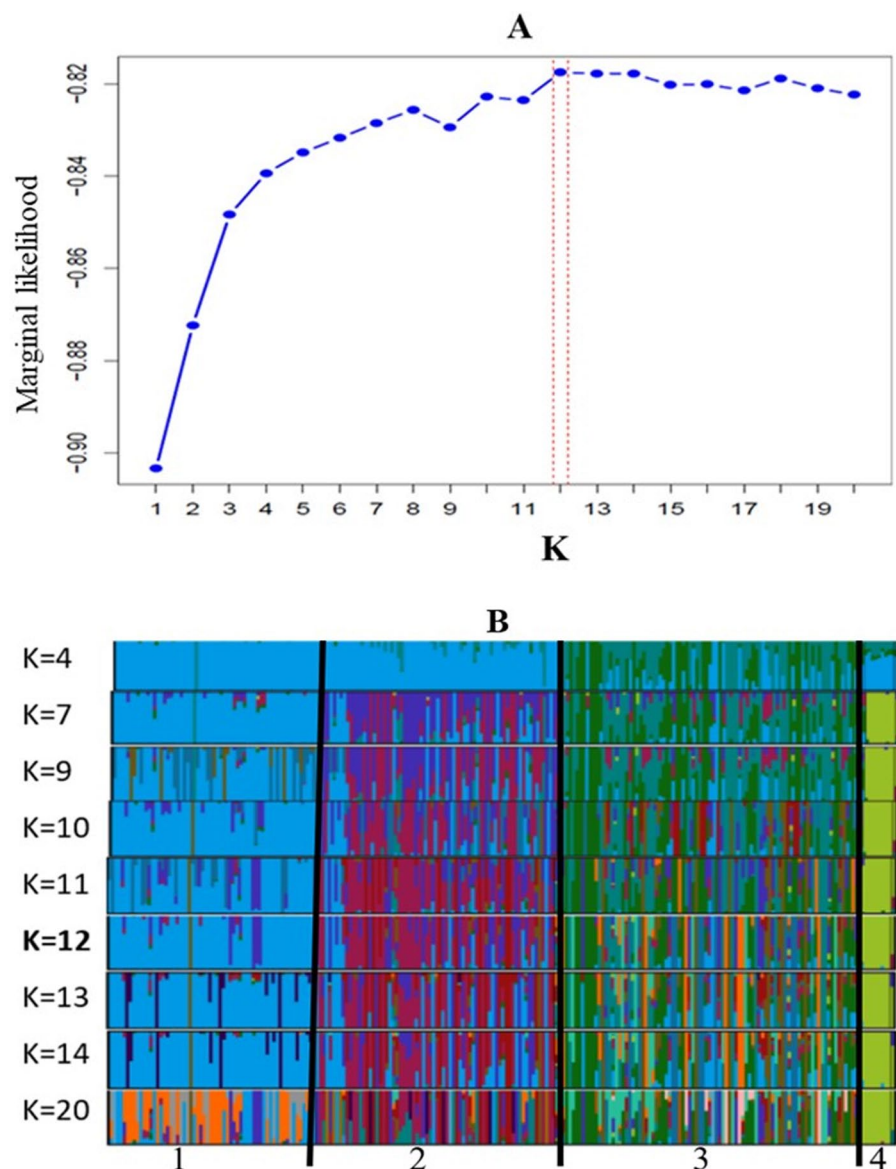
Genetic structure

The genetic structure analysis using fastSTRUCTU RE suggests that the most likely number of subpopulations is 12, i.e., the model complexity that maximizes likelihood is 12 (likelihood = −0.82) and the highest peak shows K = 12 as optimal (Fig. 2A). The results of the fastSTRUCTURE analysis are shown in Fig. 2B. The 10 wild enset accessions from South Omo make up a distinct group which is stable at all levels above K = 7. The 62 cultivated ensets from Gurage separated from Sidama and South Omo accessions and seems to make up a rather unique subpopulation, while the cultivated enset accessions from Sidama (72) and South Omo (92) represent many subpopulations.

PCA and phylogenetic relationships

Results from the principal component analysis (PCA) are presented in Fig. 3. The PCA showed that some of the populations were more clearly separated while others were clustered more closely. The first three components described 20, 18 and 9% of the total variance, respectively. PC1, with some overlap, separates Gurage accessions from accessions of the other regions, while PC2 separates the South Omo accessions in two clusters, one of them overlapping with the Sidama cluster. However, the Sidama and South Omo accessions in this cluster are partly separated by PC3, with the wild accessions clustering, as expected, with the South Omo subcluster. The phylogenetic analysis grouped the enset accessions into different clusters, to a large degree reflecting

Fig. 3 Principle component analysis (PCA) plots of PC1 and PC2 (**A**), and PC1 and PC3 (**B**). The percentages in brackets indicate the variance explained by the different PCs
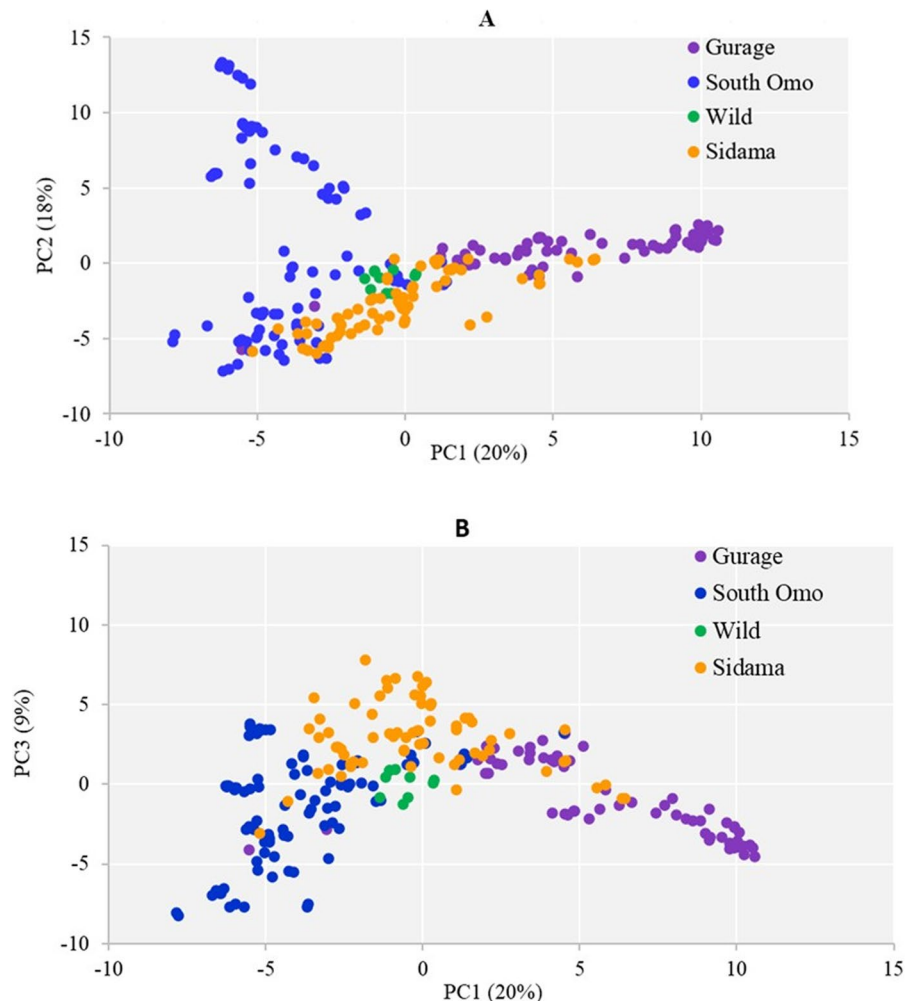
Fig. 4 Phylogenetic tree
(**A**) Maximum-likelihood
phylogenetic tree with
branch length displayed;
(**B**) Topological view of
the maximum-likelihood
phylogenetic tree. Accessions are numbered as in
Supplementary Table 1
and colored according to
their geographical origin
and cultivation status,
i.e., South Omo = blue;
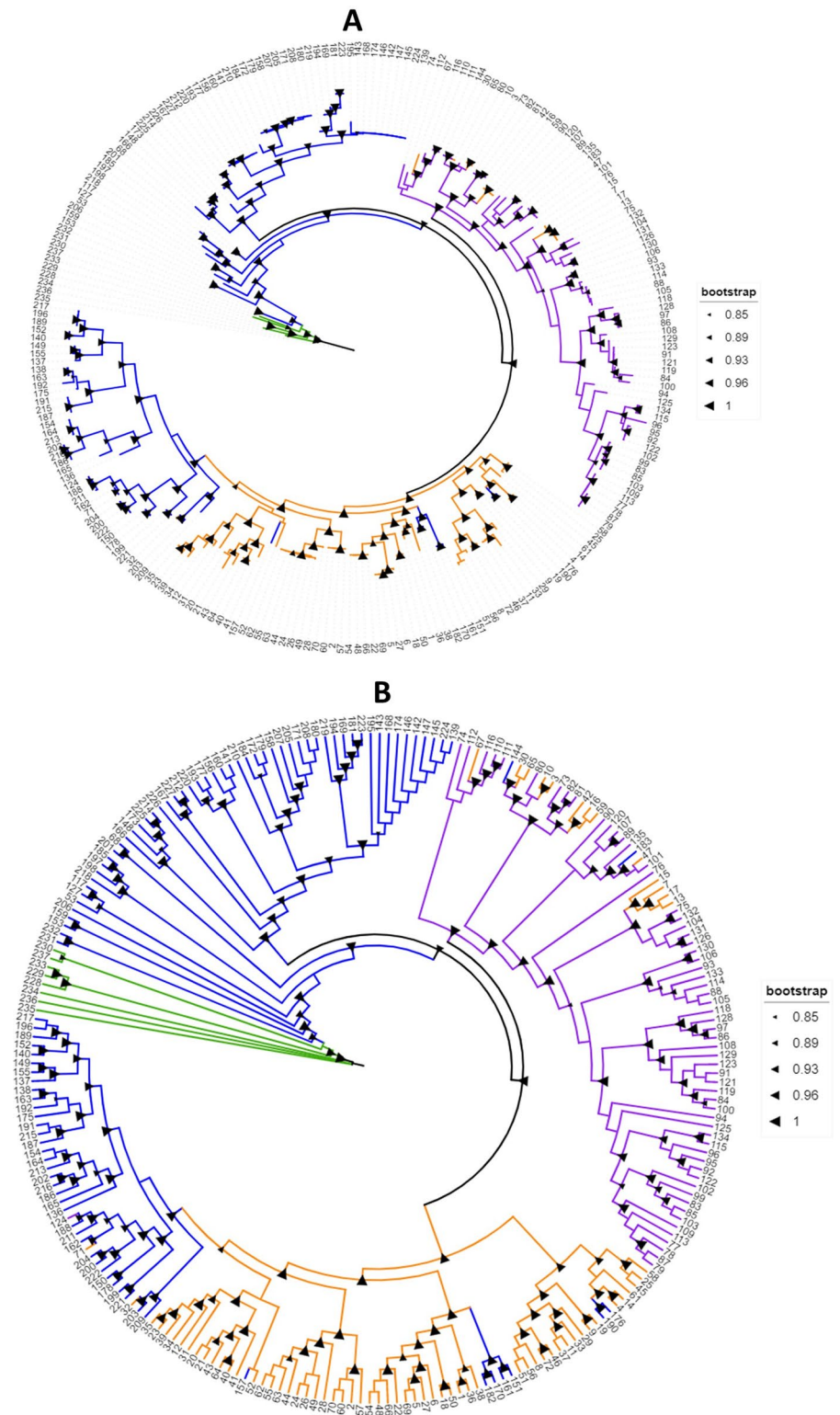Sidama = orange; Gurage = purple; wild
enset = green

Table 2 Analysis of molecular variance (AMOVA) among and within populations and regions of cultivated and wild enset

Source	df	Sum of squares	Variance component	Total variation (%)	F_{ST}
Cultivated and wild enset					
Among populations	3	16,495.01	51.40	8.79	0.09
Within populations	232	224,622.88	532.71	91.20	
Total	235	241,117.89	584.11		
Growing regions					
Among regions	2	12,822.99	44.13	7.64	0.08
Within regions	1	215,813.23	533.83	92.35	
Total	3	228,635.22	577.96		

geographical origins and cultivation status (Fig. 4). The wild accessions formed a clearly distinguished clade from the cultivated enset accessions (Fig. 4B). Generally, some accessions of cultivated enset tended to have longer branches (Fig. 4B). Interestingly, twelve accessions (19.35%) collected from the Sidama region clustered into Gurage. However, surprisingly no Gurage accessions clustered with Sidama accessions in this study. Besides, four accessions collected from Sidama and two from Gurage clustered with South Omo accessions. Some accessions have the same names in different regions, e.g., Gena, Astara and Mazia, however, they are certainly different accessions since they cluster in different clades in the phylogenetic tree (Fig. 4). The phylogenetic analysis confirms the results of the structure analysis; the most genetically unique accessions, apart from the wild accessions, are the Gurage accessions, while accessions from Sidama seems to have a mixed ancestry, whereas the South Omo accessions clearly represents two genetically diverse subgroups.

Genetic diversity and pairwise population differentiation

The results of the AMOVA analysis are presented in Table 2. Most of the genetic diversity (91.2%) is within the enset accessions, and very little (8.8%) between accessions. Analysis of the genetic differentiation between the geographic regions showed that 92.4% of the genetic diversity was within and only 7.6% between geographic regions (Table 2). Observed (H_O) heterozygosity was slightly higher in Sidama and the wild group ($H_O = 0.33$ and 0.32, respectively) than in Gurage and South Omo (both $H_o = 0.31$), while expected (H_E) heterozygosity was

Table 3 Genetic diversity among cultivated and wild enset populations based on 3505 SNP markers

	Enset populations			
	Sidama	Gurage	South Omo	Wild
Observed heterozygosity (H_O)	0.33	0.31	0.31	0.32
Expected heterozygosity (H_E)	0.32	0.30	0.32	0.36

Table 4 Average pairwise population differentiation (F_{ST})

Regions	Sidama	Gurage	South Omo
Gurage	0.07		
South Omo	0.06	0.10	
Wild	0.15	0.17	0.14

slightly lower than observed in Sidama and Gurage and higher than observed in South Omo and the wild group (Table 3). Generally, the molecular diversity was highest within the wild accessions and lowest within cultivated enset from Gurage. Pairwise population differentiation (F_{ST}) showed, as expected, that the largest subpopulation division is between the wild accession group and the cultivated (moderate to large differentiation, $F_{ST} = 0.14–0.17$), with the largest differentiation between the wild and the Gurage group. Between the cultivated enset groups, the largest differentiation is between Gurage and South Omo ($F_{ST} = 0.10$), while differentiation between Sidama and the other two groups are smaller ($F_{ST} = 0.06–0.07$) (Table 4).

Fig. 5 Candidate loci under selection were identified using F_{ST} based outlier approach (hierarchical structure model using Arlequin 3.5). F_{ST}: locus –specific genetic divergence among the populations; heterozygosity: measure of heterozygosity per locus. Loci significant at the 1% level are indicted by red dots
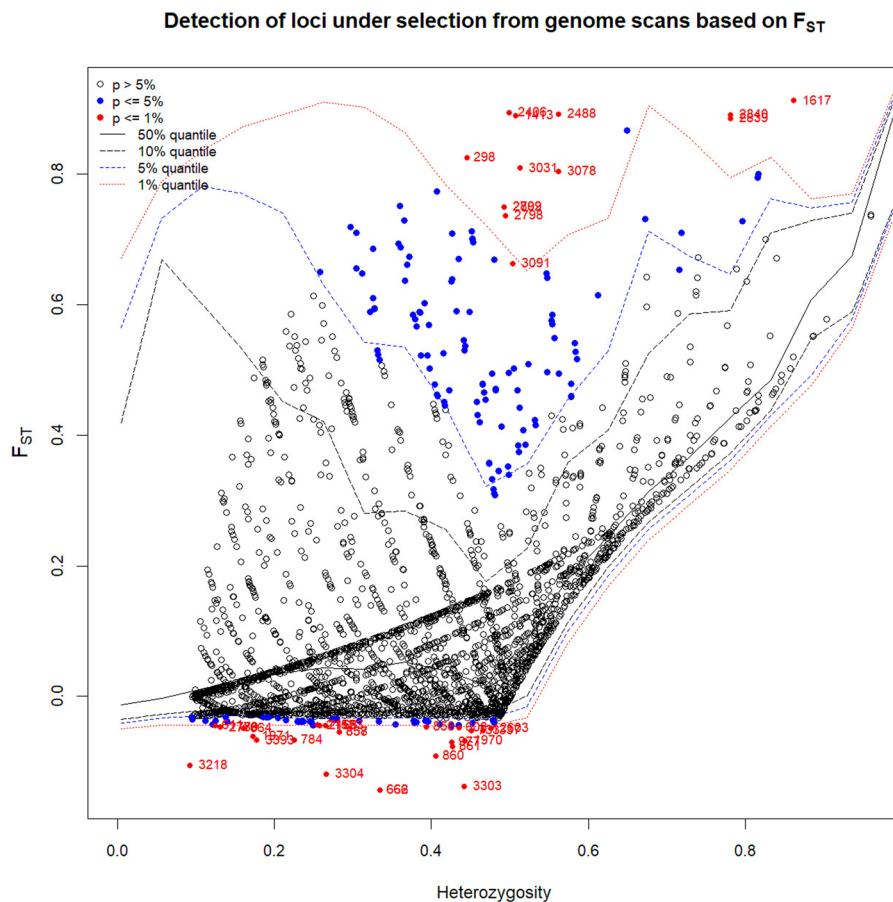

Detection of loci under selection from genome scans based on F_{ST}

Table 5 Functional annotation of outlier single nucleotide polymorphisms (SNPs) potentially involved in domestication of enset

SNP ID	Gene name	Gene function	References
E-2488	SAUR-like auxin-responsive protein family	This family plays important roles in flowering, plant growth and regulation of plant architecture	https://www.uniprot.org/uniprot/Q3E901
E-3078	Isoflavone synthase gene	Controls isoflavone accumulation and is most expressed in the developing seed	https://www.uniprot.org/uniprot/Q9M6D6
E-298	DNA binding with one finger (Dof) proteins	Shift in flowering time in the landraces through regulation of *CONSTANS*	https://www.uniprot.org/uniprot/F4IJM6
E-1617	Serine/threonine-protein kinase	Regulates flowering time by modulating the photoperiod pathway	https://www.uniprot.org/uniprot/Q944Q0
E-3031	Histone acetyltransferase	Plays a critical role in floral meristem development	https://www.uniprot.org/uniprot/Q9AR19
E-3091	R2R3-MYB transcription factor	Regulate plant growth and development	https://www.uniprot.org/uniprot/Q9SAM2

Signatures of selection and functional analyses

Signatures of balancing and directional selection were identified at 35 loci among cultivated and wild accessions using the hierarchical method (Excoffier et al. 2009) (Fig. 5). Putative balancing selection was detected at 23 loci and directional selection was detected at 12 loci (Fig. 5). Among the 12 loci, six loci have putative gene functions, while the other six loci have unknown gene functions. Putative functions of these six loci are described in Table 5.

Discussion

Population structure and differentiation between wild and cultivated Enset

In this study, a high-throughput sequencing technology was used to explore genetic diversity, population structure, and selection signatures in cultivated and wild enset accessions collected across the center of origin and domestication in Ethiopia. The ancestral admixture and phylogenetic analyses showed a clear separation between wild and cultivated enset (Fig. 2A, 4). Most probably this separation between wild and cultivated enset populations can be attributed to the difference in propagation methods (Birmeta et al. 2004; Gerura et al. 2019; Olango et al. 2015; Tesfamicael et al. 2020). It is interesting to note that cultivated enset accessions collected from regions where wild enset grows showed higher admixture and weaker clustering than those collected from regions where wild enset does not grow. This could be due to higher enset diversity in that specific region with wild enset and indicates exchange of genetic material by crossing between cultivated and wild enset. Besides, the phylogenetic tree analysis showed that populations from adjacent regions like Sidama and Gurage formed a polyphyletic group, which was not the case with distantly located populations e.g., populations from Sidama and South Omo (Fig. 4B). This genetic structure could be explained by a combination of genetic drift locally and the founder population. However, the analyses showed admixture of very few accessions irrespective of their origins whether the accessions were located isolated far apart or close like Sidama and Gurage (Fig. 4).

Remarkably, some accessions collected from Sidama clustered with Gurage accessions, suggesting that these accessions are most possibly of Gurage origin. However, no Gurage accessions clustered with accessions from Sidama. In addition, four accessions from Sidama and two from Gurage clustered with South Omo accessions. Taken together, this indicate human sharing and exchange of some clonal materials among and within regions (Gerura et al. 2019; Getachew et al. 2014). As pointed out earlier, some accessions have the same vernacular names in different regions, e.g., Gena, Astara and Mazia. However, they are certainly different accessions genetically based on their SNP profiles, and they have not been exchanged by humans even if they have the same vernacular name.

The phylogenetic tree showed long branches for the wild population from South Omo and for a few cultivated enset accessions too (Figs. 4A, 4B), suggesting high rates of nucleotide substitution and consequently high diversity. Furthermore, the phylogenetic tree revealed a relatively close association between South Omo and Sidama enset populations (Figs. 4A, 4B) and lowest F_{ST} value was found between these two populations (Table 4). Thus showing that Sidama and South Omo populations have close relationship with each other which might be due to possible vicariant evolutionary event from a single common ancestor through the fragmentation of their common ancestor's range or historical relationship (Schaal et al. 1998).

Values of the fixation index (F_{ST}) above 0.15 indicate significant differentiation between populations (Frankham et al. 2002). In this study, we observed that significant divergence between enset populations. The wild population showed moderate to large genetic differentiation from the cultivated populations from the regions, while there was relatively small differentiation between the cultivated populations. Cultivated enset is only propagated vegetatively and farmers harvest enset before seed set, while wild enset are propagated exclusively by sexual reproduction (Birmeta et al. 2004; Brandt et al. 1997). As a result of this, gene flow between cultivated and wild enset is probably very limited. Besides, the natural distribution of wild enset, as well as the farming and management practices of cultivated accessions have an impact (Birmeta et al. 2004; Olango et al. 2015). Further, limited exchange of genetic material by humans or natural factors may be considered as the main reasons

for the larger genetic differentiation observed between wild and cultivation populations (Birmeta et al. 2004; Gerura et al. 2019; Tesfamicael et al. 2020).

Importantly, moderate genetic differentiation was found between wild and cultivated enset from South Omo. This might be due to the co-existence of wild and cultivated enset in the South Omo region, where farmers introduce wild accessions into the cultivation areas and hence genetic exchange occurs between cultivated and wild populations of enset in this particular region (Shigeta 1992). In contrast, the highest F_{ST} value (0.17) was observed between wild enset and accessions from Gurage (Table 4). This shows that these accessions are more isolated from one another; most likely there is no wild enset growing in the Gurage region. Similar result can be seen from the population structure and phylogenetic analyses. The Gurage accessions are separated and formed a single cluster on their own far from the wild enset cluster (Fig. 3, 4). Another reason is that Gurage maybe has a different cultural and ethnic origin. This indicates that there is unique genetic diversity within the Gurage accessions, which is not related to the geographical distance to the other regions investigated in our study. Besides, Sidama and wild enset populations showed higher differentiation from one another. Most probably accessions from the regions are not currently breeding with one another and there is no sharing of planting materials. Concerning cultivated enset, accessions from Gurage and South Omo show low connectivity (Fig. 3, 4). This might be due to a distinct genetic profile within Gurage and South Omo accessions and possibly no frequent exchange of accessions between the two regions. Our SNPs data indicate that the cultivated and wild enset accessions are very divergent. Besides, the principal component and phylogenetic tree analysis grouped the 236 enset accessions into four major clusters, where the wild individuals clustered separately. Other enset diversity research has also reported a high level of genetic differentiation between cultivated and wild enset accessions (Birmeta et al. 2004; Gerura et al. 2019; Olango et al. 2015; Tobiaw and Bekele 2011). Also, geographic form of genetic structure was observed with consistent distinct grouping of cultivated enset accessions from Sidama, Gurage and South Omo. This knowledge of population structure and genetic diversity between cultivated and wild enset accessions is crucial for future research and breeding for new introductions.

Genetic diversity within and across populations

The large regional variation in agroecological conditions, different cultures and management relatively large geographic distances between the different enset growing regions within the country should result in large genetic diversity among regions. However, multiple lines of evidence show that the level of genetic diversity among regions (geographical areas) is low. For instance, AFLP analysis of 192 enset accessions from six growing regions showed a limited proportion of diversity among growing regions (11–13%), but a considerable diversity within regions (87–89%) (Tesfamicael et al. 2020). Earlier research also found limited diversity among growing regions compared to within regions, i.e., 13% using AFLPs (Tesfamicael et al. 2020), 4.8% using AFLPs (Negash et al. 2002) and 16% using SSR (Olango et al. 2015).These values indicated that the high proportion of genetic diversity within regions is a general feature of the enset species.

In the current study, the low genetic structuring among regions of enset that were observed both by the average pairwise F_{ST} values and AMOVA indicate that allele sharing between regions is high. The AMOVA analysis showed that the level of genetic diversity among regions is limited (7.6%) and very high within regions (92.4%) (Table 2). This is also evident from the low F_{ST} values observed between the cultivated enset accessions from the different growing regions (Table 4). These results show that genetic diversity in enset accessions are less affected by the region of origin (Schaal et al. 1998), but has rather been shaped by a long history of extensive human exchange of clonal materials among regions, and different communities may select different sources of the germplasm to suit their specific cultural needs (Gerura et al. 2019; Getachew et al. 2014; Negash et al. 2002). Furthermore, there has also most probably been extensive exchanges of clones particularly between highland and lowland regions because farmers in the latter area believe that suckers imported from the mountain areas grow better than those raised locally (Tesfaye and Lüdders 2003). Because of the large genetic diversity among accessions within regions, clonal selection based on desirable traits

may be effective for most of the natural populations in Ethiopia. In addition, the large genetic diversity within regions may be partly explained by gene flow and common origin of the populations. According to some investigations, large genetic diversity within populations is not necessarily caused by environmental heterogeneity, but could be due to historical patterns of relationship (Schaal et al. 1998).

In the present study, 3,505 SNPs markers which were polymorphic among 236 (226 cultivated and 10 wild) enset accessions were detected. This number of SNPs might be considered low relative to the 5,011 SNPs detected from 141 (120 cultivated and 21 wild enset) studied by (Tesfamicael et al. 2020). Moreover, the observed heterozygosity (H_O) and expected heterozygosity (H_E) is low (Table 3) compared to research using other DNA marker systems such as ISSR (Getachew et al. 2014) and SSRs (Gerura et al. 2019; Olango et al. 2015), but higher than with AFLP markers, which revealed lower observed and expected heterozygosity in cultivated and in wild enset populations (Tesfamicael et al. 2020). However, it is difficult to make direct comparisons between previous and the present study, due to differences in the number and types of the studied enset accessions and different SNP calling and filtering parameters applied. The reasons for the relatively low number of SNP markers detected in the present study could be frequent vegetative propagation and sharing of clones among farmers, which will reduce polymorphism. SNPs were filtered across cultivated and wild accessions; thus, the total number depends on sites that are polymorphic in the cultivated enset. If additional wild enset accessions or cultivated enset from other enset growing regions had been included in the study, the number of SNPs would probably have been higher. To rule out technical artifacts in SNP calling, we tried both the STACKS and TASSEL GBS methods for SNP calling and both derived a low number of SNPs. Also different molecular markers have different properties and will scan different regions of the genome Karp et al. (1997).

In this context, it is interesting that wild enset had lower levels of heterozygosity than expected, indicating that wild enset is a sexually propagated plant within a restricted area, which will limit gene flow and lead to inbreeding and increased homozygosity (Table 3) (Birmeta et al. 2004; Shigeta 1992). Moreover, the suitable habitats for wild enset has

been sharply declining in Ethiopia because of population growth and deforestation, and the geographical range of wild enset is more limited, possibly due to more specific ecological requirements or alternatively loss of habitat (Birmeta et al. 2004; Borrell et al. 2019; Olango et al. 2015). This reduction in effective population size might have contributed to the observed lower heterozygosity in wild enset due to the increased chances of inbreeding. This differs from what has been reported based on SSR markers (Olango et al. 2015).

However, relatively high levels of heterozygosity were observed in all cultivated populations (Table 3), which is consistent with the outcrossing nature of enset during sexual reproduction (Brandt et al. 1997; Olango et al. 2015). Enset might have improved phenotypes through heterosis, so that growers favor heterozygous cultivars in the course of selective propagation practices (Oztolan-Erol et al. 2021). Further, the current levels of enset diversity reflect frozen variation; that is diversity that arose through sexual reproduction in an ancestral population (Chapman et al. 2000). In addition, occasional gene flow from wild enset and possibly from other enset species can occur too (Birmeta et al. 2004). Other possible causes of this type of clonal diversity might be somatic mutations, introduction of new variation from outside of the cultivated populations, and introduction of new landraces from other regions (Shigeta 1990; Tsykun et al. 2017). Another possible cause might be the perennial and highly clonally propagated species that are highly selected for adaptability and productivity under cultivation, and different pollination mechanisms (Birmeta et al. 2004; Negash et al. 2002; Yemataw et al. 2016). According to Shank (1994) considerable clonal diversity is present within enset for characters associated with growth and adaptation.

Above and beyond, the highlands of southern Ethiopia form the geographical center of enset cultivation (Vavilov 1997). According to Harlan (1951), high altitude areas have high concentrations of diverse and unique landraces, and can be designated as microcenters of enset diversity. All such factors in combination or alone have resulted in a high degree of genetic diversity in the presently studied enset accessions. The most important point is that most likely differences in genetic diversity among regions are important for farmers; different accessions contribute to the high diversity that is observed at each site and

provides strong evidence for selection by humans. Enset diversity in Ethiopia may thus be extensive but it is not effectively utilized, as the available germplasm is poorly known (Borrell et al. 2019).

Genetic signatures for differential selection between cultivated and wild Enset

Little is known about the genetic makeup and population differentiation between cultivated and wild enset. Knowledge about the genetic adaptation of enset is essential for breeding strategies. A central aim of evolutionary biology is to understand the molecular basis for adaptive differences between populations (Lotterhos and Whitlock 2014). Higher genetic population differentiation for adaptive SNP than neutral SNP is expected if adaptation to local environments is the principal source of genetic differentiation (De Villemereuil and Gaggiotti 2015). F_{ST} outlier approaches has been applied to many crops, such as tomato (Sim et al. 2011), perennial ryegrass (Kovi et al. 2015), soybean (Li et al. 2014), European beech (Cuervo-Alarcon et al. 2018), banana (Hinge et al. 2022) and common bean (Papa et al. 2007) for identifying adaptive differentiation. Markers detected in these crops have been mapped to the genomic regions with known QTL/genes related to domestication.

Wild enset propagates by seed under natural condition, while cultivated enset is propagated only vegetatively by local farmers (Borrell et al. 2019; Brandt et al. 1997; Shigeta 1992). Most probably the genetic differences between wild and cultivated enset populations can be attributed to the different reproduction systems (Birmeta et al. 2004; Gerura et al. 2019; Olango et al. 2015; Tesfamicael et al. 2020). Continued vegetative propagation during cultivation can lead to loss of sexual reproduction capacity (Denham et al. 2020), thus flowering, seed development, seed size, numbers of viable seeds per fruit and per infructescence are important traits that differentiate cultivated and wild enset (Borrell et al. 2019; Brandt et al. 1997; Hildebrand 2001).

In the present study, we identified 12 candidate loci putatively under positive selection based on F_{ST} values displaying differentiation higher than the 99% limit of the confidence interval (Fig. 5, Table 5). Among them, six loci, i.e., E-2488, E-3078, E-298, E-1617, E-3031 and E-3091, might

be under direct selection. SNP annotation showed that the putative functions of these candidate loci (Table 5) are involved in different biological processes, including sexual reproduction and flowering signaling in plants, which are key players in domestication and adaptation (Borrell et al. 2019). E-2488 was identified as a SAUR-like auxin-responsive protein. Small auxin-upregulated RNAs (SAURs) is the largest family of early auxin responsive genes in higher plants regulating a wide range of cellular, physiological, and developmental processes (Ren and Gray 2015; Zhang et al. 2021). Most of the *SAUR* genes, which are part of auxin response factors (ARF) regulate cell elongation, at least in the seedlings (Sun et al. 2016). Further, Hu et al. (2015) showed higher expression of *MaARF* genes at initial days of flowering than at later stages, suggests crucial roles of the *ARF* genes in early banana fruit development. E-3078 was identified as an isoflavone synthase gene (IFS), which plays a natural role in plant defense and root nodulation. Manipulating the expression of IFS in legumes showed improved pathogen and stress responses (Jung et al. 2000). E-298 was detected as a DNA binding with one finger (Dof) protein, which is a plant-specific transcription factor having multiple roles, such as seed maturation and germination (Ruta et al. 2020). Further, Dof proteins are involved in the growth and development of banana reproductive organs (Dong et al. 2016; Venkatesh and Park 2015). E-1617 was identified as a serine/threonine-protein kinase (STK). STKs are involved in various developmental processes like cell proliferation, modification of cell shape and apoptosis. Proteomic research in somatic embryo development in banana, showed that serine/threonine-protein kinase (spot 17) was found to be highly expressed in mature somatic embryos and these proteins are associated with pattern formation and tissue specification during embryonic developmental process (Kumaravel et al. 2020). E-3031 was identified as histone acetyltransferases (HATs), which plays critical roles in the regulation of chromatin structure and gene expression. Genetic analysis and cytological study revealed that the double mutation induced severe defects in the formation of male and female gametophyte, resulting in an arrest of mitotic cell cycle at early stages of gametogenesis (Latrasse et al. 2008), thus showing their crucial roles in cell division. The final SNP, E-3091

was associated with R2R3-MYB transcription factor. These transcription factors have been shown to play regulatory roles during plant development, and responses to biotic and abiotic stress in banana (Pucker et al. 2020). Further, MYB genes *MaMYB4*, an R2R3-MYB Repressor transcription factor, negatively regulates the biosynthesis of anthocyanin in banana (Deng et al. 2021) and also *MaMYB3* is involved in fruit ripening through modulation of starch degradation (Fan et al. 2018). Moreover, two of the genes, serine/threonine-protein kinase and MYB transcription factor identified in our study were also detected recently in a similar study of enset (Tesfamicael et al. 2020).

Conclusion

Our study on enset detected a significant subdivision between cultivated and wild enset and a large molecular diversity within populations, indicating a heterogeneous collection of enset from Ethiopia. Most of the molecular diversity exists within geographical regions and very little between regions. Enset from Sidama and South Omo are more genetically diverse than enset from Gurage. Furthermore, we identified six genes involved in sexual reproduction and flowering signalling being differentially selected between cultivated and wild enset. These novel findings are useful for the conservation of genetic resources, especially under global climate changes, and contribute to the potential discovery of functional genes and genetic mechanisms related to adaptability of enset to local climatic conditions, especially drought. This is encouraging for the potential of diversifying crops also in regions where enset is not traditionally grown, such as the food insecure dry north.

Declarations

References

Abebe T (2005) Diversity in homegarden agroforestry systems of Southern Ethiopia. Dissertation, Wageningen University, The Netherlands. Wageningen University

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet 17(2):81

Baker R, Simmonds N (1953) The genus Ensete in Africa. Kew Bull 8(3):405–416

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proc R Soc Lond Ser B: Biol Sci 263(1377):1619–1626

Bekele E, Shigeta M (2011) Phylogenetic relationships between *Ensete* and *Musa* species as revealed by the trnT trnF region of cpDNA. Genet Resour Crop Evol 58(2):259–269

Bezuneh T, Feleke A, Bayie R (1967) The cultivation of the genus *Ensete* in Ethiopia. In Soil Crop Sci Soc Fla Proc, vol 27, pp 133–141

Birmeta G, Nybom H, Bekele E (2004) Distinction between wild and cultivated enset (*Ensete ventricosum*) gene pools in Ethiopia using RAPD markers. Hereditas 140(2):139–148

Biswas MK, Darbar JN, Borrell JS, Bagchi M, Biswas D, Nuraga GW, Demissew S, Wilkin P, Schwarzacher T, Heslop-Harrison JS (2020) The landscape of microsatellites in the enset (*Ensete ventricosum*) genome and web-based marker resource development. Sci Rep 10(1):15312

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120

Borrell JS, Biswas MK, Goodwin M, Blomme G, Schwarzacher T, Heslop-Harrison JSP, Wendawek AM, Berhanu A, Kallow S, Janssens S, Molla EL, Davis AP, Woldeyes F, Willis K, Demissew S, Wilkin P (2019) Enset in Ethiopia: a poorly characterized but resilient starch staple. Ann Bot 123(5):747–766

Borrell JS, Goodwin M, Blomme G, Jacobsen K, Wendawek AM, Gashu D, Lulekal E, Asfaw Z, Demissew S, Wilkin P (2020) Enset-based agricultural systems in Ethiopia: a systematic review of production trends, agronomy, processing and the wider food security applications of a neglected banana relative. Plants People Planet 2(3):212–228

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23(19):2633–2635

Brandt S, Spring A, Hiebsch C, McCabe J, Tabogie E, Diro M, Wolde-Michael G, Yntiso G, Shigeta M, Tesfaye S (1997) The "Tree Against Hunger": Enset-based agricultural systems in Ethiopia, vol 56. American Association for the Advancement of Science, Washington, DC, USA

Chapman HM, Parh D, Oraguzie N (2000) Genetic structure and colonizing success of a clonal, weedy species, *Pilosella officinarum* (*Asteraceae*). Heredity 84(4):401–409

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18):3674–3676

Costa MGDA (1984) The Itinerario of Jeronimo Lobo. Hakluyt Society, London

Cuervo-Alarcon L, Arend M, Müller M, Sperisen C, Finkeldey R, Krutovsky KV (2018) Genetic variation and signatures of natural selection in populations of European beech (*Fagus sylvatica* L.) along precipitation gradients. Tree Genet Genom 14(6):84

De Villemereuil P, Gaggiotti OE (2015) A new FST-based method to uncover local adaptation using environmental variables. Methods Ecol Evol 6(11):1248–1258

Deng G-M, Zhang S, Yang Q-S, Gao H-J, Sheng O, Bi F-C, Li C-Y, Dong T, Yi G-J, He W-D (2021) MaMYB4, an R2R3-MYB repressor transcription factor, negatively regulates the biosynthesis of anthocyanin in banana. Front Plant Sci 11:2023

Denham T, Barton H, Castillo C, Crowther A, Dotte-Sarout E, Florin SA, Pritchard J, Barron A, Zhang Y, Fuller DQ (2020) The domestication syndrome in vegetatively propagated field crops. Ann Bot 125(4):581–597

Dong C, Hu H, Xie J (2016) Genome-wide analysis of the DNA-binding with one zinc finger (Dof) transcription factor family in bananas. Genome 59(12):1085–1100

Engels J, Hawkes (1991) The Ethiopian gene centre and its genetic diversity. In: Plant genetic resources of Ethiopia. Cambridge University Press, pp 23–41

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. Heredity 103(4):285–298

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resourc 10(3):564–567

Fan ZQ, Ba LJ, Shan W, Xiao YY, Lu WJ, Kuang JF, Chen JY (2018) A banana R2R3-MYB transcription factor MaMYB3 is involved in fruit ripening through modulation of starch degradation by repressing starch degradation-related genes and MabHLH6. Plant J 96(6):1191–1205

Frankham R, Briscoe DA, Ballou JD (2002) Introduction to conservation genetics. Cambridge University Press

Gerura FN, Meressa BH, Martina K, Tesfaye A, Olango TM, Nasser Y (2019) Genetic diversity and population structure of enset (*Ensete ventricosum* (Welw.) Cheesman) landraces of Gurage zone, Ethiopia. Genet Resourc Crop Evol 66(8):1813–1824

Getachew S, Mekbib F, Admassu B, Kelemu S, Kidane S, Negisho K, Djikeng A, Nzuki I (2014) A look into genetic diversity of enset (*Ensete ventricosum* (Welw.) Cheesman) using transferable microsatellite sequences of banana in Ethiopia. J Crop Improv 28(2):159–183

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59(3):307–321

Harlan JR (1951) Anatomy of gene centers. Am Nat 85(821):97–103

Hartl DL, Clark AG, Clark AG (1997) Principles of population genetics, vol 116, 3rd edn. Sinauer Associates, Sunderland

Hildebrand E (2001) Morphological characterization of domestic vs. forest-growing Ensete 599 ventricosum (Velw.) Cheesman, Musaceae, in Sheko district, Bench-Maji Zone, ventricosum (Velw.) Cheesman, Musaceae, in Sheko district, Bench-Maji Zone, southwest Ethiopia Biologiske Skrifter (Denmark)

Hinge VR, Shaikh IM, Chavhan RL, Deshmukh AS, Shelake RM, Ghuge SA, Dethe AM, Suprasanna P, Kadam US (2022) Assessment of genetic diversity and volatile content of commercially grown banana (*Musa* spp.) cultivars. Sci Rep 12(1):1–16

Hinze LL, Hulse-Kemp AM, Wilson IW, Zhu Q-H, Llewellyn DJ, Taylor JM, Spriggs A, Fang DD, Ulloa M, Burke JJ (2017) Diversity analysis of cotton (*Gossypium hirsutum* L.) germplasm using the CottonSNP63K Array. BMC Plant Biol 17(1):37

Hu W, Zuo J, Hou X, Yan Y, Wei Y, Liu J, Li M, Xu B, Jin Z (2015) The auxin response factor gene family in banana: genome-wide identification and expression analyses

during development, ripening, and abiotic stress. Front Plant Sci 6:742

Jung W, Yu O, Lau S-MC, O'Keefe DP, Odell J, Fader G, McGonigle B (2000) Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes. Nat Biotechnol 18(2):208–212

Karp A, Edwards KJ, Bruford M, Funk S, Vosman B, Morgante M, Seberg O, Kremer A, Boursot P, Arctander P, Tautz D, Hewitt GM (1997) Molecular technologies for biodiversity evaluation: opportunities and challenges. Nat Biotechnol 15(7):625–628

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 15(5):1179–1191

Kovi MR, Fjellheim S, Sandve SR, Larsen A, Rudi H, Asp T, Kent MP, Rognli OAN (2015) Population structure, genetic variation, and linkage disequilibrium in perennial ryegrass populations divergently selected for freezing tolerance. Front Plant Sci 6:929

Kumaravel M, Uma S, Backiyarani S, Saraswathi MS (2020) Proteomic analysis of somatic embryo development in *Musa* spp. cv. Grand Naine (AAA). Sci Rep 10(1):4501

Latrasse D, Benhamed M, Henry Y, Domenichini S, Kim W, Zhou D-X, Delarue M (2008) The MYST histone acetyltransferases are essential for gametophyte development in *Arabidopsis*. BMC Plant Biol 8(1):121

Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:256–259

Li YL, Liu JX (2018) StructureSelector: a web-based software to select and visualize the optimal number of clusters using multiple methods. Mol Ecol Resour 18(1):176–177

Li Y-H, Reif JC, Jackson SA, Ma Y-S, Chang R-Z, Qiu L-J (2014) Detecting SNPs underlying domestication-related traits in soybean. BMC Plant Biol 14(1):1–8

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Mol Ecol 23(9):2178–2192

McKey D, Elias M, Pujol B, Duputié A (2010) The evolutionary ecology of clonally propagated domesticated plants. New Phytol 186(2):318–332

Negash A, Tsegaye A, van Treuren R, Visser B (2002) AFLP analysis of enset clonal diversity in south and southwestern Ethiopia for conservation. Crop Sci 42(4):1105–1111

Nuraga GW, Feyissa T, Tesfaye K, Biswas MK, Schwarzacher T, Borrell JS, Wilkin P, Demissew S, Tadele Z, Heslop-Harrison JS (2022) The genetic diversity of enset (*Ensete ventricosum*) landraces used in traditional medicine is similar to the diversity found in non-medicinal landraces. Front Plant Sci 12:3166

Olango TM, Tesfaye B, Pagnotta MA, Pè ME, Catellani M (2015) Development of SSR markers and genetic diversity analysis in enset (*Ensete ventricosum* (Welw.) Cheesman), an orphan food security crop from Southern Ethiopia. BMC Genet 16(1):98

Oztolan-Erol N, Helmstetter AJ, Inan A, Buggs RJ, Lucas SJ (2021) Unraveling genetic diversity amongst European hazelnut (*Corylus avellana* L.) varieties in Turkey. Front Plant Sci 12:1250

Papa R, Bellucci E, Rossi M, Leonardi S, Rau D, Gepts P, Nanni L, Attene G (2007) Tagging the signatures of domestication in common bean (*Phaseolus vulgaris*) by means of pooled DNA samples. Ann Bot 100(5):1039–1051

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7(5):e37135

Pino Del Carpio D, Basnet RK, De Vos RC, Maliepaard C, Visser R, Bonnema G (2011) The patterns of population differentiation in a *Brassica rapa* core collection. Theor Appl Genet 122(6):1105–1118

Pucker B, Pandey A, Weisshaar B, Stracke R (2020) The R2R3-MYB gene family in banana (*Musa acuminata*): Genome-wide identification, classification and expression patterns. PLoS ONE 15(10):e0239275

Purseglove JW (1985) Tropical crops: monocotyledons, 5th edn. Longman, London

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. Gnetics 197(2):573–589

Ren H, Gray WM (2015) SAUR proteins as effectors of hormonal and environmental signals in plant growth. Mol Plant 8(8):1153–1164

Rochette NC, Rivera-Colón AG, Catchen JM (2019) Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol Ecol 28(21):4737–4754

Rogstad SH (1992) Saturated NaCl-CTAB solution as a means of field preservation of leaves for DNA analyses. Taxon 41:701–708

Ruta V, Longo C, Lepri A, De Angelis V, Occhigrossi S, Costantino P, Vittorioso P (2020) The DOF transcription factors in seed and seedling development. Plants 9(2):218

Schaal B, Hayworth D, Olsen KM, Rauscher J, Smith W (1998) Phylogeographic studies in plants: problems and prospects. Mol Ecol 7(4):465–474

Shank R (1994) The enset culture: a technical report on *Enset ventricosum* or 'False Banana.' United Nations-Emergencies Unit for Ethiopia, Addis Ababa

Shigeta M (1990) Folk in-situ conservation of ensete (*Ensete ventricosum* (WEL W.) E.E. Cheesman): towards the interpretation of indigenous agricultural science of the Ari. southwestern Ethiopia. Afr Study Monogr 10(3):93–107

Shigeta M (1992) The ethnobotanical study of ensete (*Ensete ventricosum*) in the southwestern Ethiopia. Kyoto University, Japan

Sim S, Robbins M, Van Deynze A, Michel A, Francis D (2011) Population structure and genetic differentiation associated with breeding history and selection in tomato (*Solanum lycopersicum* L.). Heredity 106(6):927–935

Simmonds NW (1962) The evolution of the bananas. Longmans, Green, London

Spring A, Haile B, Tesfaye S, Abebe Y, Amaldegn A, Wolde-Michae G, Tabogie E, Surur O, Tsegaye A, Shimeles S (1996) Enset farming systems in southern region, Ethiopia: Report on a rapid rural appraisal in Gurage, Hadiya,

and Sidama zones Deutsche Gesellschaft for Technische Zusammenarbeit (GTZ), Addis Ababa, Ethiopia

Sun N, Wang J, Gao Z, Dong J, He H, Terzaghi W, Wei N, Deng XW, Chen H (2016) Arabidopsis SAURs are critical for differential light regulation of the development of various organs. Proc Natl Acad Sci 113(21):6071–6076

Tesfamicael KG, Gebre E, March TJ, Sznajder B, Mather DE, Lopez CMR (2020) Accumulation of mutations in genes associated with sexual reproduction contributed to the domestication of a vegetatively propagated staple crop, enset. J bioRxiv

Tesfaye B, Lüdders P (2003) Diversity and distribution patterns of enset landraces in Sidama, Southern Ethiopia. Genet Resour Crop Evol 50(4):359–371

Tobiaw DC, Bekele E (2011) Analysis of genetic diversity among cultivated enset (*Ensete ventricosum*) populations from Essera and Kefficho, southwestern part of Ethiopia using inter simple sequence repeats (ISSRs) marker. Afr J Biotech 10(70):15697–15709

Tsegaye A, Struik PC (2001) Enset (*Ensete ventricosum* (Welw.) Cheesman) kocho yield under different crop establishment methods as compared to yields of other carbohydrate-rich food crops. Netherl J Agric Sci 49(1):81–94

Tsegaye A, Struik PG (2002) Analysis of enset (*Ensete ventricosum*) indigenous production methods and farm-based biodiversity in major enset-growing regions of southern Ethiopia. Exp Agric 38(3):291–315

Tsegaye A (2002) On indigenous production, genetic diversity and crop ecology of enset (*Ensete ventricosum* (Welw.) Cheesman). Wageningen University and Research.

Tsykun T, Rellstab C, Dutech C, Sipos G, Prospero S (2017) Comparative assessment of SSR and SNP markers for inferring the population genetic structure of the common fungus *Armillaria cepistipes*. Heredity 119(5):371–380

Vavilov NI (1951) The origin, variation, immunity and breeding of cultivated plants, vol 72, no 6. LWW, p 482

Vavilov N (1997) Five continents (eds) LE Rodin. Semyon Reznik and Paul Stapleton, International Board for Plant Genetic Resources, Rome, Italy, pp 1–197

Venkatesh J, Park SW (2015) Genome-wide analysis and expression profiling of DNA-binding with one zinc finger (Dof) transcription factor family in potato. Plant Physiol Biochem 94:73–85

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38(6):1358–1370

Westphal E, Stevels JMC, Stevels JMC (1975) Agricultural systems in Ethiopia, vol 826. Centre for Agricultural Publishing and Documentation, Wageningen

Wright S (1978) Variability within and among populations. Evolution and the Genetics of Populations

Yemataw Z, Tesfaye K, Taddese T, Dejene T, Muzemil S, Shashego Z, Ambachew D (2016) Genetic variation for corm yield and other traits in Ethiopian enset (*Ensete ventricosum* (Welw.) Cheesman). J Plant Breed Crop Sci 8(9):150–156

Zeberga A, Yemataw Z, Sinebo W, Musemil S (2014) On farm cultivar diversity of enset (*Ensete ventricosum* W.) in Southern Ethiopia. J Agric Dev 4:62–85

Zhang H, Yu Z, Yao X, Chen J, Chen X, Zhou H, Lou Y, Ming F, Jin Y (2021) Genome-wide identification and characterization of small auxin-up RNA (SAUR) gene family in plants: evolution and expression profiles during normal growth and stress response. BMC Plant Biol 21(1):1–14