Maintenance and expansion of genetic and trait variation following domestication in a clonal crop

Oliver W. White[1], Manosh Biswas[2], Wendawek M. Abebe[3], Yann Dussert[4], Firew Kebede[3], Richard A. Nichols[4], Richard J. A. Buggs[1,4], Sebsebe Demissew[5], Feleke Woldeyes[6], Alexander S.T. Papadopulos[7], Trude Schwarzacher[2,8], Pat J.S. Heslop-Harrison[2,8], Paul Wilkin[1], James S. Borrell[1]

Corresponding author: J. S. Borrell, j.borrell@kew.org

Addresses

[1] Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, UK

[2] Department of Genetics and Genome Biology, University of Leicester, LR1 7RH, UK

[3] Department of Biology, Hawassa University, Hawassa, Ethiopia

[4] School of Biological and Behavioural Sciences, Queen Mary University of London, London

[5] Department of Biology, Addis Ababa University, Addis Ababa, P.O. Box 3293,5 Ethiopia

[6] Ethiopian Biodiversity Institute, Addis Ababa, Ethiopia

[7] Schools of Natural Sciences, University of Bangor, Gwynedd, LL57 2DG, UK

[8] Key Laboratory of Plant Resources Conservation and Sustainable Utilization/Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China

Orchid IDs

Oliver White	0000-0002-7151-723X
Manosh Biswas	0000-0001-9228-1927
Wendawek M. Abebe	
Yann Dussert	0000-0003-1900-7715
Firew Kebede	
Richard Nichols	0000-0002-4801-9312
Richard J. A. Buggs	0000-0003-4495-3738
Sebsebe Demissew	
Feleke Woldeyes	
Alexander S.T. Papadopulos	0000-0001-6589-754X
Trude Schwarzacher	0000-0001-8310-5489
Pat J.S. Heslop-Harrison	0000-0002-3105-2167
Paul Wilkin	0000-0003-4982-7175
James S. Borrell	0000-0001-9902-7681

Abstract

Clonal propagation enables favourable crop genotypes to be rapidly selected and multiplied. However, the absence of sexual propagation can lead to low genetic diversity and accumulation of deleterious mutations, which may eventually render crops less resilient to pathogens or environmental change. To better understand this trade-off, we characterise the domestication and contemporary genetic diversity of Enset (*Ensete ventricosum*), an indigenous African relative of bananas (*Musa*) and principal starch staple for 20 million Ethiopians. Wild enset is strictly sexually outcrossing, but in cultivation is exclusively propagated clonally and associated with diversification and specialisation into hundreds of named landraces. We applied tGBS sequencing to generate genome-wide genotypes for 192 accessions from across enset's cultivated distribution, and surveyed 1340 farmers on enset agronomic traits. Overall, reduced heterozygosity in the domesticated lineage was consistent with a domestication bottleneck that retained 37% of wild diversity. However, excess low frequency putatively deleterious missense mutations present as heterozygotes suggested accumulation of mutational load in clonal domesticated lineages. Our evidence indicates that the major domesticated lineages initially arose through historic sexual recombination associated with a domestication bottleneck, followed by amplification of favourable genotypes through an extended period of clonal propagation. Among domesticated lineages we found significant phylogenetic signal for multiple farmer identified food, nutrition and disease resistance traits and little evidence of contemporary recombination. Development of future-climate adapted genotypes may require crop breeding, but outcrossing risks exposing deleterious alleles as homozygotes. This trade-off may partly explain the ubiquity and persistence of clonal propagation over recent centuries of comparative climate stability.

Keywords

Agrobiodiversity, Domestication, *Ensete ventricosum*, Food security, Clonal evolution, Ethiopia, tGBS, Genotyping-by-sequencing

Abstract (Amharic translation)

አጽህሮተ-መጣጥፍ

ኢ-ጾታዊ-መራቦ ተመራጭ ባሂሪን የሚያስከትሉ የሰብል ይዘተ-በራሂዎች በተፋጠነ ሁኔታ እየተለዩ እንዲባዙ ያደርጋል። ይሁን እንጂ የጾታዊ መራቦ አለመኖር የበራሂ ተለያይነት መመናመንና ጎጂ የሆነ በራሂያዊ-ቅየራን ሊያበራክት ይችላል። ይህ ሁኔታ ደግሞ የሰብል ተክሎችን በሽታንም ሆነ የአየር ንብረት ለውጥን የመቋቋም ባህሪ ሊቀንስ ይችላል። በእነዚህ አሉታዊና አዎንታዊ ገጽታዎች ረገድ ያለውን ሁኔታ በተሻለ መገንዘብን ባለመ ሁኔታ ከስራስር ሰብሎች አንዱ በሆነው የእንሰት ተክል ተላምዶና በራሂያዊ ተለያይነት ላይ ያተኮረ የባህሪ ትንተና ጥናት ተከናውኗል። እንሰት በአፍሪካ ከሚገኙ የሙዝ ዝሪያዎች ወገን የሚመደብ ሲሆን ከ20 ሚሊዮን ለሚበልጡ ኢትዮጵያውያን አብይ የታይል ሰጪ ምግብ ምንጭ ነው። እንሰት ከእርሻ ሰብለነት ባለፈ በዱር በቀልነትም ይገኛል። ዱር በቀሉ እንሰት የማይበላ ሆኖ የመራቢያ ዘዴው ጾታዊ ነው። በሰብልነት የሚለማው እንሰት በጾታዊ መንገድ ሊራባ ቢችልም አርሶ አደሮች ኢ-ጾታዊ-መራቦን በመጠቀም ያባዙታል። በዚህም መንገድ በመቶዎች የሚቆጠሩ የተለየ ባህሪና ስያሜ ያላቸው አይነቴዎች እንዲበለጽጉ ሆኗል። በዚህ ጥናት ቲ.ጂ.ቢ.ኤስ. (tGBS) ሲኩዌንሲነግ የተባለ ዘዴን በመጠቀም የ192 አይነቴዎችን ይዘተ-በራሂዎች የማንበብ ሥራ የተከናወነ ሲሆን 1340 አርሶ አደሮችን በማሳተፍ በአይነቴዎቹ ባህሪ ዙሪ መረጃ ተሰብስቧል። በሰብልነት በሚለማው የእንሰት ወገን የድቅል በራሂያዊነት ይዘት አነስተኛ ሲሆን ይህም በማላማዱ ሂደት የተከሰተው በራሂያዊ ደፈቃ ከዱር በቀሉ ብዝሀነት 37% ተለያይነት ብቻ እንዲወረስ ካሳረፈው ተጽእኖ ጋር የሚቆራኝ ነው። ይሁን እንጂ አልፎ አልፎ ጎጂ የሆነ በራሂያዊ-ቅየራ በድቅል-በራሂ መልክ የሚከሰት መሆኑ በኢ-ጾታዊ መንገድ በሚራባው የተላመደው የእንሰት ወገን ላይ በጊዜ ሂደት የበራሂያዊ-ቅየራ ጫና እየበረታ መሄዱን ያመለክታል። በአጠቃላይ የጥናት ውጤቱ እንደሚያሳየው ዋና ዋናዎቹ የእንሰት የዘር ሀረጎች የተከሰቱት ቀደም ሲል በነበረ በራሂያዊ-ብወዛ፣ በማላማዱ ሂደት በተከሰተ በራሂያዊ ደፈቃና ይህን ተከትሎ በተፈጠረ የተለያይነት መመናመን፣ አንዲሁም በዘመናትነ የማላመድ ሂደት ጠቃሚነት ያላቸውን ይዘተ-በራሂዎች ከማበራከት ጋር በተያያዘ በተደረጉ መረጣዎች ምክንያት ነው። ከተላመዱ የእንሰት አይነቴዎች አንጻር ጠንካራ የምግብነት፣ የንጥረ-ምግብ እና በሽታን የመቋቋም ስርወዘራዊ የባህሪ መመሳሰል የተስተዋለ ቢሆንም በመካከላቸ ወቅታዊ በራሂያዊ-ብወዛ ስለመኖሩ

የሚያሳይ የጠነከረ መረጃ አልተገኘም። መጻኢው የብዛዘር ጥበቃና አስተዳደር የአየር ንብረት ለውጥን ከመቋቋም አካያ በጸታዊ-መራቦ መንገድ የሰብል ህብረዘር ማጎልበትን ተፈላጊ ሊያደርግ የሚችልበት ሁኔታ ቢኖርም ይህ የተዋልዶ ዘዴ መሰል-በራሂን በማበራከት አሉታዊ ተጽእኖ ሊያሳርፍ እንደሚችል ይታሰባል። ይህም እውነታ በእንሰት ልማት የኢ-ጸታዊ-መራቦ ዘዴ ተዘውታሪና ብቸኛ አማራጭ ስለመሆኑ መሰረታዊ ምክንያት ተደርጎ ሊወሰድ ይችላል

Introduction

More than half of domesticated food crop species are clonally propagated (Meyer, Duval, & Jensen, 2012). Clonal propagation is advantageous to farmers because it allows rapid multiplication of agronomically useful genotypes, without the need for many generations of breeding to fix desirable traits (Denham et al., 2020; Mckey, Elias, Pujol, & Duputié, 2010). However due to the absence of recombination, the effects of selection on a single locus extends to more linked sites, leading to reduced genetic diversity (Mckey et al., 2010) and impeding the removal of deleterious mutations from a population (i.e. Hill-Robertson Interference; Charlesworth, Betancourt, Kaiser, & Gordo, 2009; Chen et al., 2019; Ramu et al., 2017). Over time clonal propagation could erode genetic potential for farmer-led landrace and trait diversification, as well as adaptation to environmental change and emerging pests and pathogens (Dodd & Douhovnikoff, 2016). These processes may particularly impact tropical agricultural systems where clonally propagated crops are disproportionately abundant (Denham et al., 2020).

To better understand the trade-offs associated with clonal propagation, we investigate the domestication and diversification of the major Ethiopian food security crop enset (*Ensete ventricosum* (Welw.) Cheesman). Enset, known colloquially as Ethiopia's "tree against hunger" (Brandt et al., 1997), is a giant monocarpic herb from the same family as the bananas (Musaceae), whose pseudostems and corms (fleshy tissue from the stem and roots) provides a starch staple for 20 million people (Borrell et al., 2020). Wild enset reproduces sexually and does not produce lateral rhizome suckers (shoots arising from the roots underground) unlike bananas. In contrast, domesticated enset is exclusively clonally propagated via human-mediated removal of the apical meristem (Borrell et al., 2020), triggering sucker production. Furthermore, enset is harvested for the pseudostem and underground corm prior to flowering, strongly limiting the potential for outcrossing in cultivation. As a result, wild and domesticated enset display a clear and analytically tractable distinction between sexual and clonal modes of evolution (Tamrat et al., 2022; Tesfamicael et al.,

2020). The extensive indigenous knowledge associated with clonal propagation methods, compared to the absence of knowledge of sexual propagation, suggests that this cultivation system has been ubiquitous for a considerable length of time (Borrell et al., 2019).

Despite reduced recombination and a cultivated distribution restricted to South-Western Ethiopia, domesticated enset is remarkably diverse, with >1,200 named landraces occurring across a wider range of environments that any other local crop (Supporting Information Table S 1). This diversity includes extensive variation, including plant morphology (Yemataw et al., 2017), food and nutrition traits (Borrell et al., 2020; Tamrat et al., 2020), fibre quality (Blomme et al., 2018), medicinal value (Yemataw, Tesfaye, Zeberga, & Blomme, 2016), pest (Kidane, Meressa, Haukeland, Hvoslef-Eide, & Coyne, 2021) and disease tolerance (Hunduma, Kassahun, Hilu, E., & Oli, 2015).

Genetic variation in clonally propagated crops commonly originates from sexual populations before and during domestication (Myles et al., 2011). This includes secondary sexual contact with wild populations, for example integration of volunteer seedlings in cassava (Sardos et al., 2008) and ennoblement in yams (Scarcelli et al., 2006). In such scenarios, extant genetic variation is influenced by the genetic architecture of useful traits and the strength of farmer selection through historic domestication bottlenecks, often resulting in a reduction in heterozygosity (Miller & Gross, 2011; Tesfamicael et al., 2020). However, as sexual recombination becomes rare, mitotic processes such as the accumulation of spontaneous somatic mutations (Balloux, Lehmann, & De Meeûs, 2003) may become increasingly important (Foster & Aranzana, 2018; Zhou, Massonnet, Sanjak, Cantu, & Gaut, 2017). We expect novel variation from these sources to accumulate over time and be detectable as increased heterozygosity in clonal enset lineages. Whilst these could be involved in driving novel traits (e.g. pineapple; Chen et al., 2019), the reduced efficiency of natural selection may result in increased mutational load.

In this study, we genotype 192 enset landraces from across the cultivated distribution to investigate the genetic basis for clonal crop domestication and diversification. We first characterise the

differentiation of domesticated enset from partly sympatric wild populations and test for phylogenetic signal in farmer-reported agronomic traits. Second, to evaluate the ubiquity of clonal propagation as the principal reproduction strategy, we used D-statistics to evaluate the frequency of sexual reproduction across wild and domesticated lineages. Third, we assess evidence for a reduction in diversity associated with a domestication bottleneck, as well as characterising novel diversity that has arisen subsequently. We use our findings to better understand this trade-off and the consequences for indigenous clonal agricultural systems under major future climate and food security challenges.

Methods

Collection of enset landraces

We collected leaf samples of 225 domesticated enset individuals across the zone of enset cultivation in 2017-19. These represented 192 named landraces identified by farmers, and were selected to encompass the wide phenotypic and vernacular landrace diversity. We also collected 14 wild *E. ventricosum* accessions from populations in western Ethiopia and seven putatively semi-domesticated accessions, occurring on the periphery of the domesticated enset growing regions. Semi-domesticated individuals were identified by farmers as not belonging to either domesticated enset or wild populations, and as such farmers did not attribute them to a named landrace (Haile & Tesfaye, 2022). They are generally recorded near cultivated plots, were not planted by farmers and are occasionally used for food, but otherwise their origins are uncertain. Finally, we included one horticultural accession (*Ensete ventricosum* 'Maurelii') and three outgroup samples comprising closely related African *Ensete* species from the Living Collection at RBGKew (Figure 1A; Supporting Information Table S 2).

Sequencing and variant calling

DNA was isolated from silica-dried leaf material using a modified CTAB-based protocol (Doyle, J., Doyle, 1987). Samples were submitted to Data2Bio (Iowa, United States) for library preparation and sequencing. We used tuneable genotyping-by-sequencing (tGBS) which offers advantages over other reduced representation approaches for having higher SNP calling accuracy at heterozygous sites and less missing data (Ott et al., 2017). DNA samples were digested with the restriction enzymes NspI and BfcCI/Sau3AI before being sequenced using an Ion Proton platform.

Raw reads were quality filtered using Trimmomatic version 0.36 (Bolger, Lohse, & Usadel, 2014), removing leading and trailing bases below quality 15 and applying a 4 bp sliding window with average quality 15. Quality filtered reads that did not have restriction site overhangs for the NspI (CATG) or BfcCI/Sau3AI (GATC) enzymes were removed using cutadapt version 4.1 (Martin, 2011). To achieve uniform read length for SNP calling, we truncated reads to 80 bp and discarded shorter reads using the process_radtags module of STACKS version 2.41 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). Retained reads were mapped to the published draft genome assembly for *E. ventricosum* landrace Bedadeti (GenBank accession GCA_000818735.3; Yemataw et al., 2018) using BWA version 0.7.17 (Li & Durbin, 2010). Reads that did not map uniquely, or were flagged as alternative (XA:Z) or chimeric (SA:Z) alignments, were removed using SAMtools version 1.9 (Li et al., 2009). A summary of the number of reads at each step is provided in Supporting Information Table S 3. We then used the reference assembly pipeline from STACKS, assembling loci and filtering SNPs using the gstacks and populations modules respectively. Loci originating from contaminant sequences and potentially paralogous loci were identified and excluded from downstream analyses (Supporting Information Methods S 1). Two variant datasets were generated for this study. The first dataset, hereafter the 'phylogenetic dataset', was used for phylogenetic analysis, identification of putative clones, phylogenetic signal in landrace traits, SNPs under selection and inference of sexual

97 recombination. Whereas the second dataset, hereafter referred to as the 'population genetic
98 dataset', was used to investigate evidence of a population bottleneck in domesticated enset.

99 For the *phylogenetic dataset*, we retained SNPs present in ≥80% of individuals and a minimum minor
100 allele count of three. We further removed putative copy number variants (CNVs) using filtering
101 procedures developed by Dorant et al. (2020) to retain only high confidence singleton SNPs. To
102 minimise the confounding influence of linkage, only the first SNP per locus was retained and all
103 remaining SNPs showing strong genotype correlation were removed.

104 For the *population genetic dataset*, we retained a single representative per multilocus genotype
105 (MLG) for wild and domesticated samples (i.e., clone correction; see Methods: Identification of
106 putative clones). Semi-domesticated and outgroup samples were excluded, resulting in a sample size
107 of 141, comprising 128 domesticated and 13 wild samples. SNP calling was repeated as above, with
108 the exception that SNPs were retained if present in 80% of all individuals, 80% of domesticated
109 samples, or 80% of wild samples. These filters were used to ensure that SNPs unique to either the
110 wild or domesticated populations were retained despite different sample sizes. SNPs were not
111 filtered using a minimum allele count as we expect a fraction of somatic mutations to be present at
112 very low frequency. In addition, we removed SNPs categorised as duplicated or diverged using the
113 filtering procedures developed by Dorant et al. (2020). In general, we were cautious in our SNP
114 filtering approach based on F_{IS}, which is expected to be negative under clonal variation due to an
115 excess of heterozygotes relative to random mating.

116 *Phylogenetic analysis of enset diversity*

117 To identify and visualise genetic clusters using the phylogenetic analysis dataset, we used principal
118 component analysis (PCA) implemented using dudi.pca in the adegenet package (Jombart, 2008;
119 Jombart & Ahmed, 2011; R Core team, 2020). For phylogenetic reconstruction we employed a
120 maximum likelihood (ML) approach implemented in RAxML Next Generation v.0.9.0 (Kozlov, Darriba,

Flouri, Morel, & Stamatakis, 2018) using a supermatrix of assembled loci concatenated into a single alignment with missing data coded as Ns. The optimal model of sequence evolution was identified using ModelTest-NG v.0.1.5 (Darriba et al., 2019) based on the corrected Akaike Information Criterion. We then generated 5,000 random and 5,000 parsimony starting trees, selecting the tree with the lowest log-likelihood, before performing 5,000 bootstrap replicates. To ensure sufficient replicates were used, we performed the bootstrapping convergence test with a cut-off threshold of 0.03. For phylogenetic trees with large sample sizes and relatively few variant patterns, it is common to find relatively low support for basal nodes based on Felsenstein's bootstrap (BS) values, which require a replicate branch to match a reference branch exactly to be accounted for in the BS value. Therefore, we employed the transfer bootstrap expectation (Lemoine et al., 2018; Lutteropp, Kozlov, & Stamatakis, 2019) approach for BS values, which is less sensitive to misplaced taxa and appropriate for a large dataset composed of numerous closely related landraces. To provide additional support, we implemented a similar approach in IQ-TREE v.1.6.12 (Nguyen, Schmidt, Von Haeseler, & Minh, 2015) (Supporting Information Methods S 2), with topologies compared using phytools and visualised using ggtree (R Core team, 2020; Yu, Smith, Zhu, Guan, & Lam, 2017), and a neighbour network using SplitsTree v.4.15.1 (Huson & Bryant, 2006).

Identification of putative clones

Individuals putatively from the same clonal lineage were identified by calculating pairwise genetic distances using bitwise.dist scaled by missing data in the R package poppr version 2.8.6 (Kamvar, Brooks, & Grünwald, 2015) and clustering samples below a given threshold into multilocus genotypes (MLGs). The threshold used to define MLGs was predicted using cutoff_predictor with the "farthest" algorithm. This method is suited to large SNP datasets, where it may not possible to define clonal individuals based solely on genetic identity due to somatic mutations and sequencing/SNP calling errors (Le Cam et al., 2019; Wang et al., 2017). We further identified MLGs with samples that were non-monophyletic on our RAXML-NG phylogenetic tree and removed these

146 from downstream analyses. To estimate the total number of MLGs in the enset growing regions we
147 used an accumulation curve approach implemented in vegan (Oksanen et al., 2019). We randomly
148 sampled genotypes across a range of sample sizes and calculated the number of MLGs. We then
149 extrapolated using the best fitting model, estimating error over 100 replicate runs.

150 *Phylogenetic signal of landrace traits*

151 During field collections 1340 farmers were asked to rank landraces present on their farm for
152 important agronomic properties including: kocho quality (the main enset derived food product),
153 bulla quality (an additional food product), medicinal value, fiber quality and enset bacterial wilt
154 susceptibility or tolerance (Supporting Information Methods S 3). The cumulative number of times a
155 landrace was observed across all farms was also recorded. After correction for the number of
156 landrace observations, we tested for evidence of a phylogenetic signal in trait scores by calculating
157 Abouheif's C_{mean} and Pagel's λ using the R package phylosignal (Keck, Rimet, Bouchez, & Franc,
158 2016). These indices differ in that they are based on principles of spatial autocorrelation and
159 evolutionary models respectively. However, both were found to be reliable measures of
160 phylogenetic signal in simulation studies (Münkemüller et al., 2012). Null traits, modelled as random
161 and under Brownian motion (BM) were added to our analysis to act as negative and positive
162 controls, respectively. To minimise the potential for farmer or researcher misidentification, perhaps
163 due to differences between ethnic groups and languages, trait values for samples were collated only
164 from survey data collected within the same administrative zone and ethnic group as for a given
165 genotyped landrace.

166 *Identifying SNPs under selection*

167 We used Bayescan version 2.1 (Foll & Gaggiotti, 2008) and an FDR threshold of 0.05 to identify sites
168 consistently differentiated between the domesticated and wild samples. For this analysis, only a
169 single representative per MLG was retained and sites with a MAF less than 0.05 were removed. For

each significant site we extracted up to 10kb up and downstream of the mapping position in the Bedadeti genome using bedtools slop version 2.28.0 (Quinlan & Hall, 2010). Specifically, we sought to test the hypothesis that wild and domesticated enset are differentiated by genes involved in flower initiation and seed development as suggested by Tesfamicael et al. (2020). Reference coding sequences (CDS) were downloaded for *Musa acuminata* subsp. *malaccensis* (v.2.0) from the Banana genome hub (Droc et al., 2013) and used to annotate regions under putative selection using blastn with an e-value of 1×10^{-30} and retaining the single best hit per CDS. To identify gene ontology (GO) terms that are enriched in regions under selection we employed topGO in R (Alexa & Rahnenfuhrer, 2021). Specifically, we compared GO terms for annotations of regions under selection against a background reference of annotations for all regions.

Frequency of sexual recombination

To identify evidence of gene flow between lineages, we first visualised haplotype sharing patterns with RADpainter and fineRADstructure (Malinsky, Trucchi, Lawson, & Falush, 2018). An individual co-ancestry matrix based on haplotype data was generated by running 100,000 MCMC generations after a burn-in of 100,000 generations. D-statistics (ABBA-BABA tests) were also computed using Dsuite (Malinsky, Matschiner, & Svardal, 2020) at both the population and individual level. D-statistics take a four taxon/population pectinate tree with the topology (((P1,P2),P3),O) and identify discordant ancestral (A) and derived (B) allele patterns denoted as ABBA or BABA. Under the null expectation, ABBA and BABA allele patterns occur due to incomplete lineage sorting and their relative frequencies are expected to be equal. However, if P3 has hybridised with either P2 or P1, we would expect an asymmetry in ABBA or BABA allele frequencies. Significance was assessed using a block-jackknife approach to calculate a *Z*-score and associated *P*-value which was adjusted for multiple tests using a Benjamini and Hochberg correction (Benjamini & Hochberg, 1995). Tests performed at the individual level were grouped by the population type of taxa occurring at the P2

and P3 position and summarised by the number and percentage of positive tests per topology. For these analyses, only a single representative per MLG was retained.

To test the possible origin of the semi-domesticated samples in our study, we simulated hybrids in adegenet (Jombart, 2008; Jombart & Ahmed, 2011). This uses the allelic frequencies of two parental populations to sample simulated gametes following a multinomial distribution. Simulated crosses were performed for F1 hybrids of domesticated × domesticated and domesticated × wild. In addition, F2 backcrosses were simulated between F1 domesticated × wild crosses and domesticated. For the simulated crosses, a random subsample of ten individuals was taken for the domesticated and wild populations, and ten hybrids were generated for each type of cross. The simulated hybrids were compared to the phylogenetic dataset using PCA implemented using dudi.pca in adegenet.

Evidence for a domestication bottleneck

Using the *population genetic dataset*, we calculated the proportion of heterozygous site per individual (excluding invariant sites). Minor allele frequencies (MAF) and the inbreeding coefficient (F_{IS}) were calculated using adegenet and hierfstat in R (Goudet, 2005; Jombart & Ahmed, 2011). Density histograms for the proportion of heterozygous sites, MAF and F_{IS} were plotted for each population. Significant differences between populations were tested using one-way ANOVA or Mann-Whitney-Wilcoxon tests for normally and non-normally distributed data respectively.

To evaluate the strength of a potential population bottleneck during enset domestication, we identified alleles that were: 1) private to the wild population (lost during domestication), 2) shared between the wild and domesticated population (retained during domestication, or introgressed after domestication) or 3) private to the domesticated population (novel in the domesticated population). For example, the proportion of wild variation retained during domestication was calculated as shared alleles / private wild alleles + shared alleles. To account for biases in sample sizes, the number of alleles in each category and proportion retained was estimated using

rarefaction over 100 iterations. Finally, SNPs across these categories were annotated with their putative effect on protein coding sequences using SnpEff version 4.3 (Cingolani et al., 2012) and the genome annotation for *E. ventricosum* landrace Bedadeti (Yemataw et al., 2018). The proportion of alleles with a given annotation type (i.e., synonymous, non-synonymous etc) was calculated. The proportion of SNPs in each category and their annotation was quantified using rarefaction as above.

Results

Sequencing and variant calling

Sequencing generated approximately 787 M reads (Supporting Information Table S 3), with a mean of 3.15 M per sample (range 0.47 – 10.93 M). Filtering low quality reads or those without restriction cut sites removed 25% of reads, retaining an average of 2.36 M per sample. Truncating the quality filtered reads to a uniform length of 80 bp and filtering those that did not meet this threshold removed approximately 30% of reads, resulting in a mean of 1.65 M reads per sample. After filtering 79.80% of retained reads mapped uniquely to the reference genome for *E. ventricosum* landrace Bedadeti (GenBank accession GCA_000818735.3; Yemataw et al., 2018). Mapped reads were assembled into 1.67M loci with a mean coverage of 15.6x per sample (range: 6.9 – 40.5). For the *population genetic dataset*, variant calling with populations resulted in 27,562 loci and 32,441 SNPs. After removing putative contaminant, paralogous sites, 9,162 loci and 12,409 SNPs were retained for downstream analysis. For the *population genetic dataset*, variant calling resulted in 27,007 loci and 70,533 SNPs, of which 24,760 loci and 64,004 SNPs were retained after the removal of putative contaminant or paralogous sites.

Phylogenetic analysis of enset diversity

Phylogenetic analyses with RAxML-NG (Figure 2A; Supporting Information Figure S 1) found clear differentiation and strong monophyletic support for domesticated (transfer bootstrap expectation bootstrap [BS]=1.00), wild (BS=0.89) and outgroup enset samples (BS=0.99). The seven semi-

domesticated enset were paraphyletic with respect to domesticated enset (BS=0.96, 0.99 and 1.00). Comparative approaches using IQ-TREE (Supporting Information Figure S 2) and neighbour networks (Figure 3A) recovered highly similar overall topologies and consistent domesticated subclades (Supporting Information Figure S 3). However, the support for early diverging branches in the domesticated clade was typically poor. This is evident when we collapse RAxML-NG branches with a bootstrap support value < 0.75, resulting in large polytomies, but numerous small, well-supported terminal clades (Supporting Information Figure S 4).

Across the domesticated enset phylogeny, we found highly significant phylogenetic signal for multiple uses and traits (Table 1). This is consistent with the observation that farmers cultivate multiple different landraces specialised for different purposes (Blomme et al., 2018; Borrell et al., 2020; Yemataw et al., 2017).

Identification of putative clones

We identified 121 domesticated enset multilocus genotypes (MLGs; Figure 3B). Histograms of pairwise genetic distance show a characteristic double peak in domesticated enset, similar to that reported in domesticated banana (Sardos, Perrier, & Dole, 2016) with the first peak putatively corresponding to variation within MLGs, and the latter to variation between MLGs. By contrast we observe only a single peak in wild and semi-domesticated accessions (Supporting Information Figure S 5). A single wild sample was identified as a putative clone by our analysis and was removed from all downstream analyses as a likely sibling. Based on the observed number of MLGs, species accumulation curve analysis estimated that the enset cultivating region contains approximately 195 MLGs if adequately sampled (CI 187-202; Figure 3C).

Identifying SNPs under selection

We identified 26 outliers across 26 tGBS loci between domesticated and wild enset from a total of 7580 sites (Supporting Information Table S 5). Of these, only 5 GO term (GO:0048364 root

development, GO:0016791 phosphatase activity, GO:0008168 methyltransferase activity, GO:0043531 ADP binding and GO:0031072 heat shock protein binding) were identified as significantly over-represented (Supporting Information Table S 6). In contrast to Tesfamicael et al. (2020), GO terms associated with flower initiation or seed development were not overrepresented. Therefore, we did not find evidence of selection against sexual reproductive capacity in clonal lineages, consistent with recent empirical studies (Tamrat et al., 2022).

Frequency of sexual recombination

Analyses of haplotype co-ancestry did not identify widespread patterns of recent hybridisation within the domesticated lineage, supporting the predominance of clonal reproduction. By contrast, wild and semi-domesticated samples were grouped together based on haplotype sharing (Figure 4A). D-statistics at the population level with the topology (((P1: domesticated, P2: semi-domesticated), P3: wild), P4: outgroup), found evidence of hybridisation (D=0.047, P <0.01). Of 551,300 tests performed between individuals, 56,683 (10.28%) were significant after Benjamini and Hochberg correction for multiple tests (Supporting Information Table S 4).

Evidence for a domestication bottleneck

We found significantly lower observed heterozygosity in domesticated enset ($F_{(1,139)}$ = 16.55, p < 0.01, mean domesticated = 0.067, mean wild = 0.086; Figure 5A), providing evidence of a domestication bottleneck. However, under clonal reproduction we might also expect spontaneous somatic mutations to accumulate over time, increasing heterozygosity. This is evidenced by significantly lower minor allele frequency in domesticated enset (W = 1.76×10^9, p < 0.01, mean domesticated=0.058, wild=0.089; Figure 5B), showing a higher frequency of rare mutations. Similarly, rarefaction analyses revealed that the number of alleles private to the wild population (lost during domestication) far exceed the number of alleles shared (retained during domestication) or private to the domesticated population (novel changes since domestication; Figure 5D). Overall, we

290 estimate that approximately 37% of wild alleles were retained (Figure 5). We found that 16% of
291 variation was unique to domesticated lineages, though we caution that our sampling of wild
292 populations may be incomplete, and some populations may have become locally extinct due to
293 extensive land use change in Ethiopia.

294 We found significantly lower F_{IS} in domesticated lineages ($W = 2.42\times10^{8}$, $p < 0.01$, mean
295 domesticated=0.080, wild=0.284; Figure 5C), indicative of a shift to clonal reproduction (Reichel,
296 Masson, Malrieu, Arnaud-Haond, & Stoeckel, 2016). The lack of sexual recombination in clonal
297 lineages results in non-random association between loci (Halkett, Simon, & Balloux, 2005). These
298 associations make the selection against deleterious alleles less efficient due to effects known
299 collectively as Hill-Robertson Interference (Charlesworth et al., 2009; Comeron, Williford, & Kliman,
300 2008). Consistent with this process we also find that alleles private to the domesticated populations
301 have a significantly higher proportion of putatively deleterious missense mutations (Figure 5E),
302 indicating that they are ineffectively purged from the genome and thus deleterious alleles are
303 accumulating in domesticated enset lineages. However, we note that these could be functionally
304 important as part of the domestication process.

305 ## Discussion

306 Domestication has transformed enset from a wild, unpalatable, outcrossing giant forest herb, to the
307 predominant staple starch crop for >20 million people. This has been achieved through the
308 combination of sexual and clonal propagation, mediated by multiple stages of human selection. We
309 hypothesise that the major domesticated lineages initially arose through historic sexual
310 recombination. The existence of an early sexually recombining domesticated population ancestral to
311 contemporary domesticated lineages is supported by the following evidence. First, we observe
312 genetic differentiation between domesticated enset and wild populations. Direct farmer selection of
313 clonal lineages with desirable attributes from wild ancestral populations (i.e., instantaneous
314 domestication) is not supported by our phylogeny, because domesticated lines are not distributed

amongst wild lineages and do not represent fixed genotypes overlapping with the sampled wild gene pool. Similarly, lower bootstrap support for early diverging branches (Figure 3A) suggests that contemporary domesticated clades result from multiple independent clonal lines selected from an ancestral population, as observed in other clonally propagated crops (Mckey et al., 2010; Scarcelli et al., 2006). An ancestral recombining population could generate conflicting topologies and reduce phylogenetic support at the base of domesticated sub-clades, consistent with observed patterns.

Second, we observe a strong bottleneck during enset domestication, as indicated by reduced heterozygosity. In a scenario of instantaneous domestication (i.e., an immediate switch to clonal propagation using favourable genotypes identified from wild populations), heterozygosity would remain unchanged (and later increase with accumulation of somatic mutations). Many perennial crops such as fruit trees display weaker bottlenecks, perhaps due to long generation times and difficulty in evaluating traits until the plant reaches maturity (Miller & Gross, 2011). Enset can be scored for non-fruiting traits during immaturity (e.g., growth rate), so performance can be assessed by farmers in as little as two to three years and selected plants propagated, potentially explaining the existence of a stronger bottleneck than in other species with similar life histories. Given that wild enset does not generate suckers or adventitious shoots, it is most plausible that initial domestication was underway prior to the cultural advent of clonal propagation methods. This theory would require fewer concurrent agricultural innovations to explain adoption of enset as a food plant (Zohary, 2004).

Consequences of clonal propagation

The ubiquity of clonal propagation in enset suggests that it has been highly advantageous in selecting favourable genotypes, however our results highlight long term risks. We demonstrate the accumulation of putative somatic mutations, illustrated by a lower frequency of minor alleles in domesticated enset (i.e., heterozygous singletons). Our analysis suggests that these are more likely to be deleterious due to a higher proportion of missense annotations, and therefore that they are

being less effectively purged. Over time, exclusive clonal propagation is therefore likely to introduce a limited amount of novel variation, as illustrated by variation within MLGs, but also increase mutational load in domesticated enset. Conversely, it is also plausible that some losses of function could be beneficial, for example by stopping the production of an unpalatable compound. Overall, the largest source of variation remains that which persists through the historic domestication bottleneck, though we identify 16% novel variation. This underlines the value of conserving wild populations which harbour a library of potentially useful genetic variation.

The rarity of hybridisation between clonal lineages in domesticated enset is achieved through agronomic practices which include the harvesting of enset prior to flowering (Borrell et al., 2020). We propose that semi-domesticated samples may represent feral landraces resulting from the breakdown of exclusive clonal propagation and introgression from wild populations (Wu, Lao, & Fan, 2021). An alternative explanation is that semi-domesticated enset are relicts from an ancestral domesticated lineage. However, a hybridization scenario seems more plausible because simulated hybrid domesticated × wild and the F1 domesticated backcrosses overlapped with the semi-domesticated samples in a principal component analysis (Figure 4C) and semi-domesticated enset are localised to the periphery of the enset growing regions where germplasm management is substantially relaxed, enabling more frequent flowering and subsequent outcrossing.

Vernacular and genetic diversity

Of the 18 landrace names for which we collected multiple samples, only one was monophyletic. While this may be partly explained by farmer misidentification during sampling, many landrace names describe phenotypic traits, and thus convergent use of these descriptors across the 64 local languages may be widespread. Indeed, some landrace names in enset (e.g. Nechwe, referring to white tissue colour) are also applied to indigenous yam (*Dioscorea*) landraces (pers. obs. J Borrell). This suggests that landrace names may not be synonyms for corresponding genotypes, outside of the local area (for this reason we restricted our vernacular and trait analyses to observations within

each administrative zone). Synonymy of cultivar names has been reported for bananas in Eastern Africa (Karamura, Karamura, & Tinzaara, 2012), and discordance in landrace identification and genetic relationships also occur in cassava (Rabbi et al., 2015) and Ethiopian wheat (Hodson et al., 2020). Consequently, our research highlights the need for a more formalised description of enset landrace diversity, as has been achieved for bananas (International Plant Genetic Resources Instit., 1996; Ruas et al., 2017) which in turn will aid experimental evaluation of landraces with agronomic traits of interest (e.g., Borrell, Gebremariam, & Wendawek, 2021; Hunduma et al., 2015; Kidane et al., 2021).

Conclusion

The ubiquity of clonal propagation in enset agriculture clearly illustrates its value to farmers in selecting and multiplying favourable genotypes – and this approach has been widely adopted in other species. Increased genetic load, which remains hidden in heterozygotes during clonal reproduction, could also be one of the factors maintaining clonal propagation as the preferred agronomic system. Whilst this strategy may have been optimal in the recent Holocene, a period of relative climate stability, we know that Ethiopia is already experiencing a warming climate, and that climate impacts are projected to become more severe throughout this century (Koch et al., 2022; McSweeney, New, & Lizcano, 2010). We suggest that for enset, and clonal crops more widely, the trade-off between clonal propagation and sexual recombination may shift further towards the latter under climate change. Due to a lack of indigenous agronomic knowledge associated with enset breeding, this is likely to require collaboration between enset farmers, who hold a significant repository of knowledge on enset traits, and researchers in a joint effort to identify or generate future-adapted enset genotypes.

Funding

This work was supported by the GCRF Foundation Awards for Global Agricultural and Food Systems Research, entitled, 'Modelling and genomics resources to enhance exploitation of the sustainable and diverse Ethiopian starch crop enset and support livelihoods' [Grant No. BB/P02307X/1], the GCRF I-FLIP grant 'Enhancing enset agriculture with mobile agri-data, knowledge interchange and climate adapted genotypes to support the Enset Center of Excellence' [BB/S018980/1] and GCRF Agrisystems award entitled, 'Landscape scale genomic environment data to enhance the food security of Ethiopian agri-systems' [BB/S014896/1]. JB was additionally supported by a Future Leader Fellowship at the Royal Botanic Gardens, Kew.

Data Accessibility and Benefit-Sharing

Raw sequence reads can be downloaded from the NCBI short read archive under BioProject PRJNA866136. Individual genotype data are available on DataDryad (to be submitted upon manuscript acceptance). Associated sample metadata including georeferences are provided in Supporting Information Table 2. Associated scripts to process data and reproduce analyses are provided here: https://github.com/o-william-white/Enset_tGBS

Benefits Generated: This paper is the product of an equitable research collaboration between scientists from Ethiopia and the UK. All collaborators are included as co-authors. Genetic material was transferred for analysis in the UK following Ethiopian material transfer laws coordinated by the Ethiopian Biodiversity Institute, underpinned by an MOU with Addis Ababa University (OERPC/657/06/14). The research addresses the primary need of researchers for enset genetic resources to accelerate research on enset agriculture. Benefits from this research accrue from the sharing of our data and results on public databases as described above.

Acknowledgements

We thank Dr Laura Kelly, Dr Rafal Gutaker, Dr Melanie-Jayne Howes and Dr Marybel Soto-Gomez for comments on earlier versions of this manuscript. We also thank Admas Berhanu, Balemso Bekele, Djene Sertse and numerous drivers and facilitators for supporting extended collaborative fieldwork in Ethiopia. Most importantly we thank the hundreds of farmers who gave us their valuable time to share their incredible knowledge of enset.

Author contributions

JB, PHH, SD and PW designed the study. JB, WA and FK performed the fieldwork collections. PHH, TS and MB coordinated and led the DNA extraction and tGBS sequencing. OW, MB and JB led the bioinformatic analyses with input from YD and RN. All authors contributed to the interpretation of the results. OW prepared the first draft and all authors contributed to the preparation of the final manuscript.

Figures

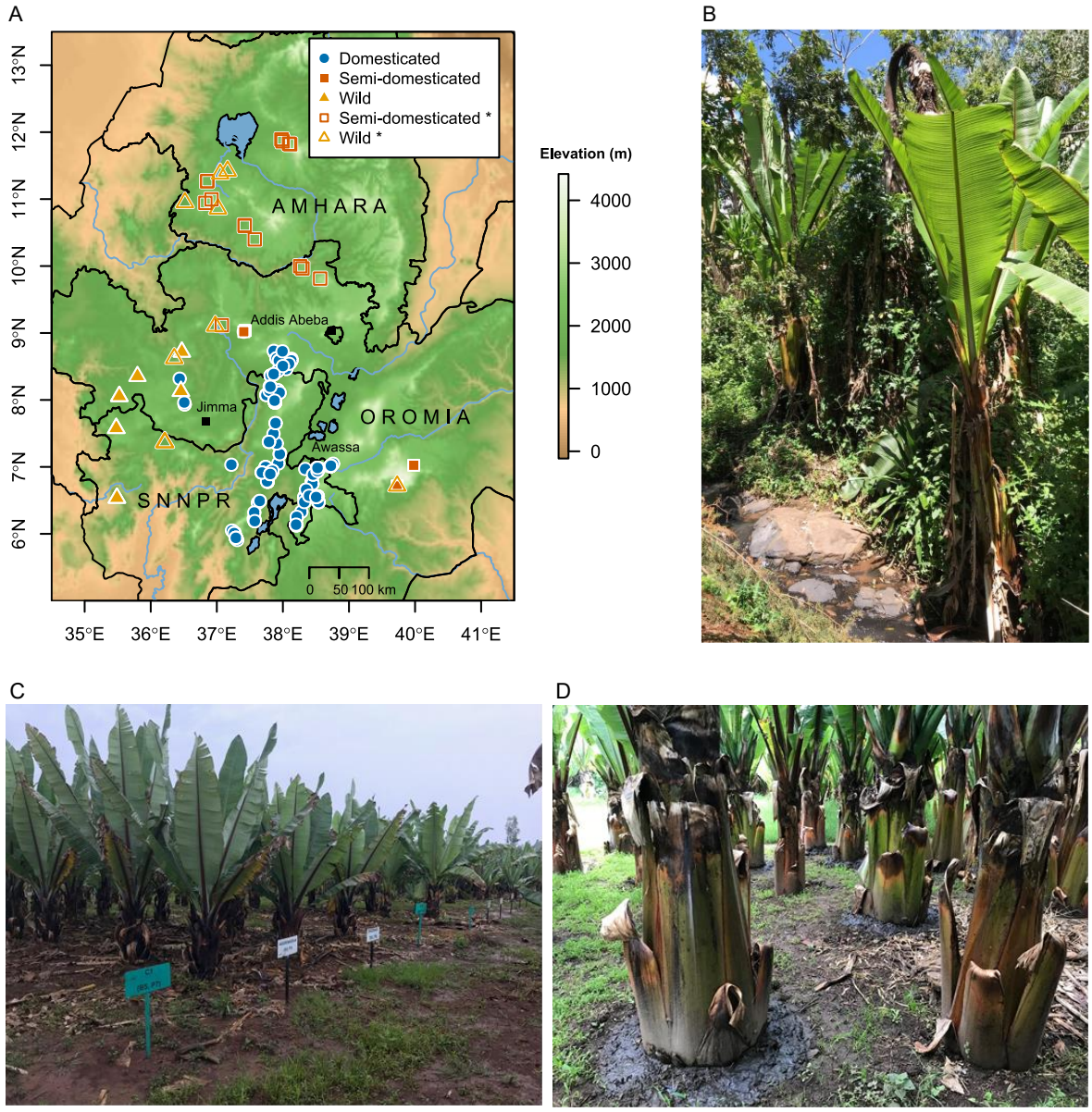

Figure 1 Distribution and phenotypic diversity of enset. (A) Distribution of domesticated, semi-domesticated and wild samples collected in Ethiopia for the present study. Points with no fill (marked with an asterisk in the legend) refer to further collection localities of wild and semi-domesticated samples not included in this study. (B) Images of wild and (C,D) domesticated enset, (Photos, J Borrell).

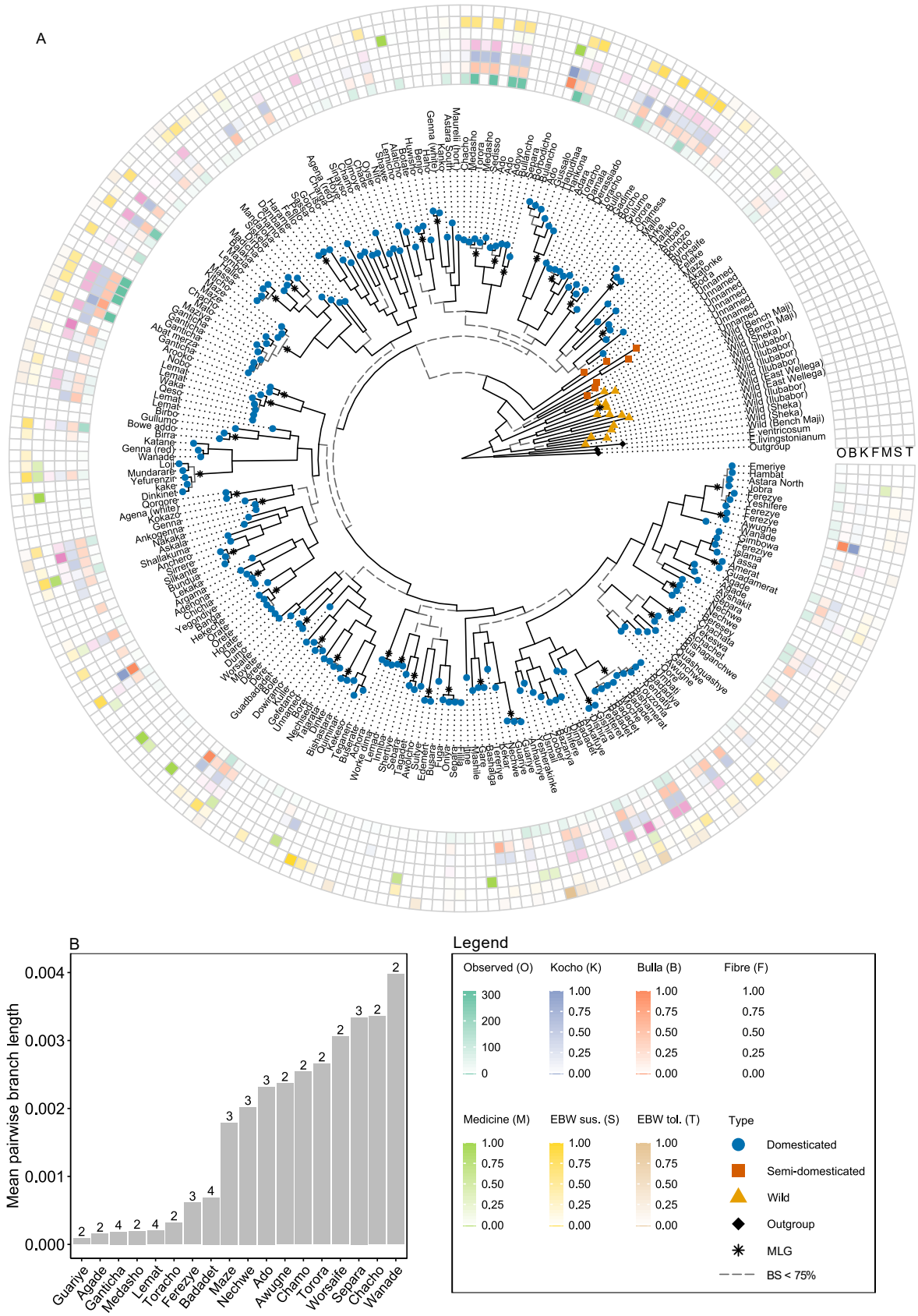
A
O B K F M S T
B
Mean pairwise branch length
0.000
0.001
0.002
0.003
0.004
Guariye
Agade
Ganticha
Medasho
Lemat
Toracho
Ferezye
Badadet
Maze
Nechwe
Ado
Awugne
Chamo
Tororа
Worsaile
Separa
Chacho
Wanade
Legend
Observed (O)
Kocho (K)
Bulla (B)
Fibre (F)
Medicine (M)
EBW sus. (S)
EBW tol. (T)
Type
Domesticated
Semi-domesticated
Wild
Outgroup
MLG
BS < 75%

Figure 2 Phylogenetic and trait diversity of enset in Ethiopia. (A) Maximum likelihood phylogeny of enset produced using RAxML-NG. Branch lengths are shown except for the outgroup taxa which were collapsed into a single node with a branch length equal to the longest branch length in domesticated and wild enset. Branches with transfer bootstrap expectation (TBE) values <75% are shown as grey dashed lines. Branches leading to multilocus genotypes (MLGs) estimated using poppr are annotated with an asterisk. Tips are coloured according to accession type (domesticated, wild or outgroup. (B) Mean pairwise branch lengths between the 18 landraces for which duplicate samples were collected with the number of replicates above each bar.

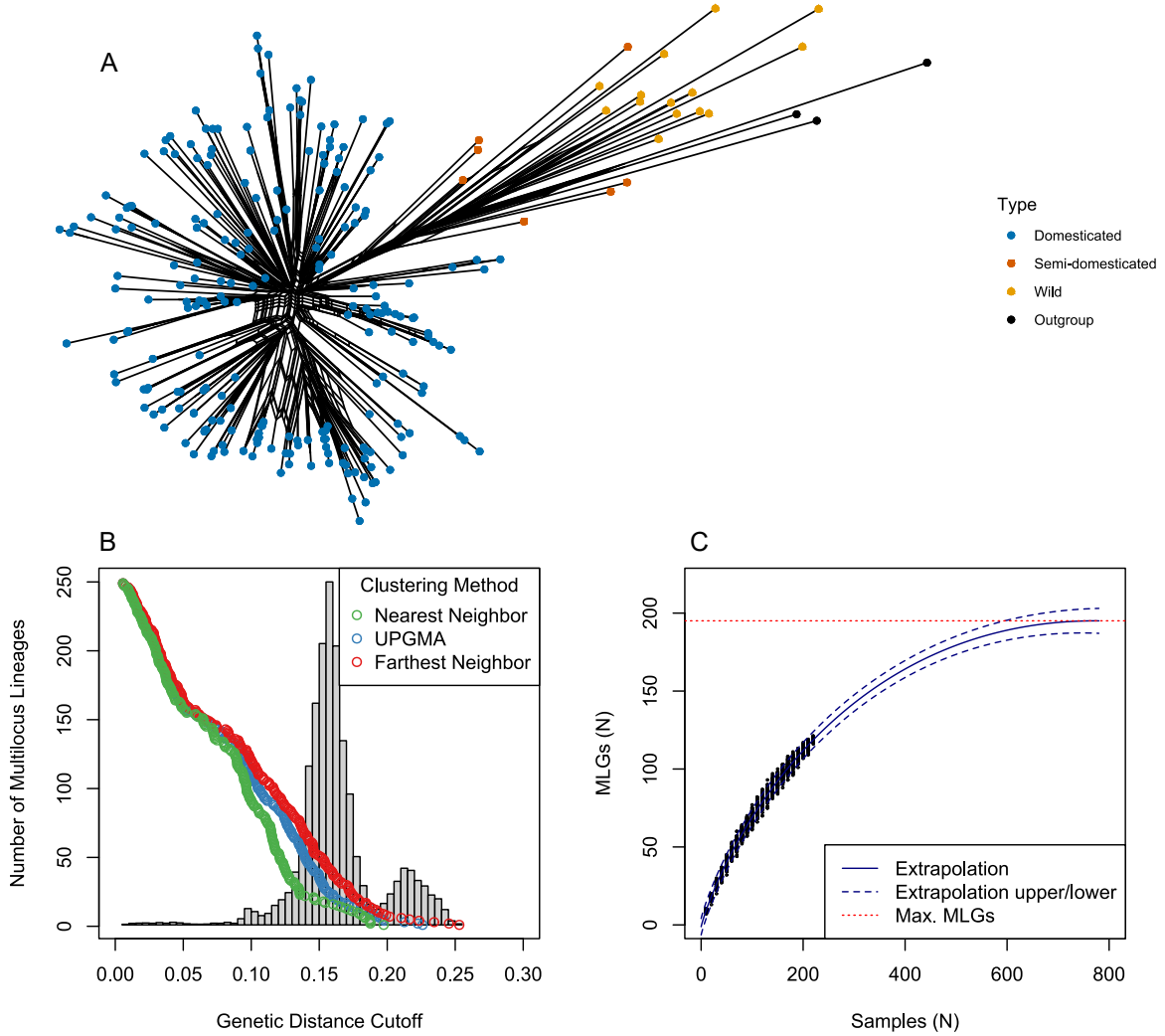

Figure 3 Genetic and multilocus genotype (MLG) diversity of enset. (A) Neighbour network of enset produced using SplitsTree with tips annotated by sample type. (B) Histogram of pairwise genetic distances with vertical lines representing thresholds for collapsing multilocus genotypes (MLGs) using nearest (0.040), UPGMA (0.051) and farthest (0.051) algorithms. The number of MLGs identified using each algorithm was 158 using the nearest and 145 for UPGMA and farthest. (C) MLG accumulation curve estimated using rarefaction with an extrapolation based on the best supported linear model. The maximum number of MLGs based on the extrapolation is shown.

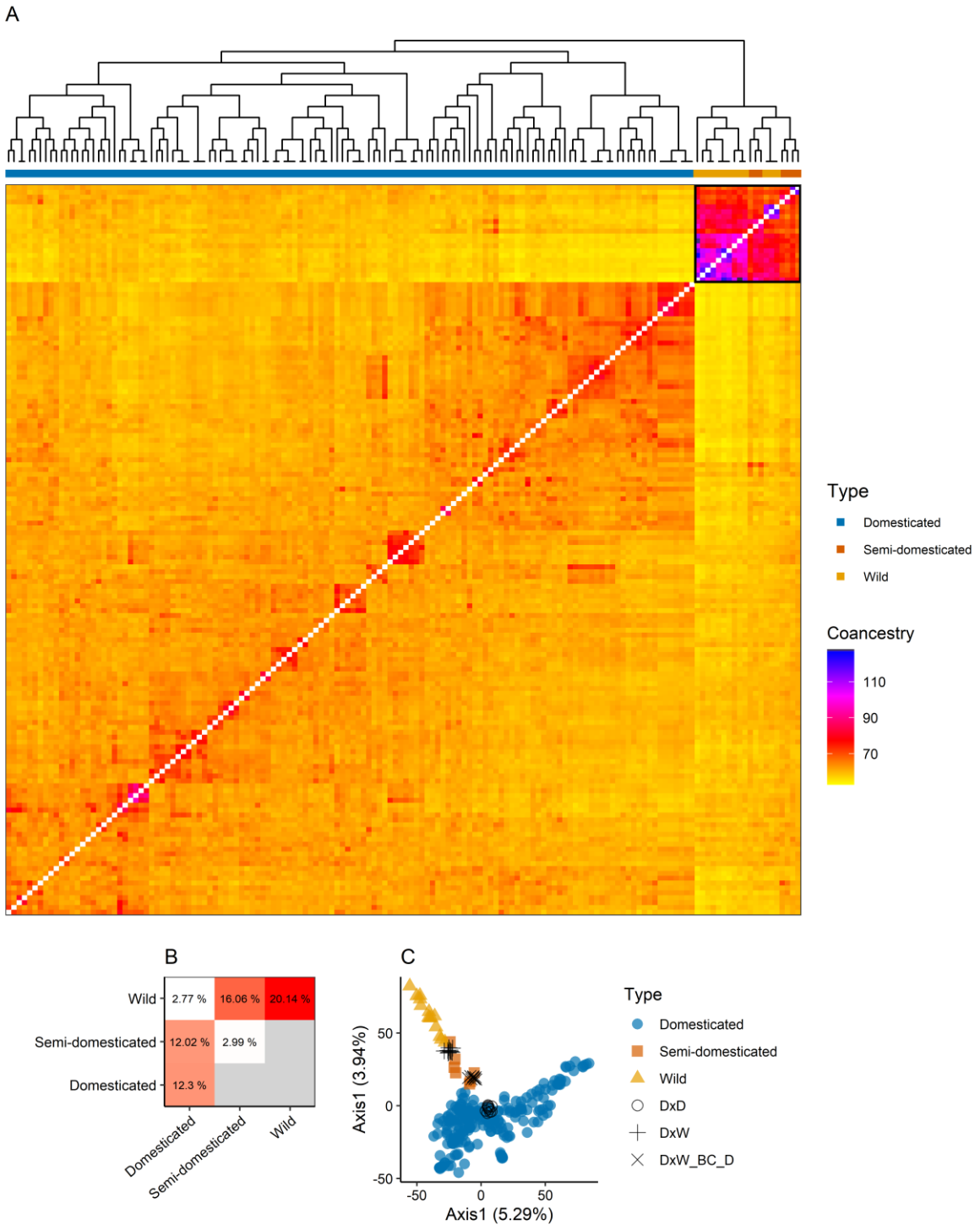

Figure 4 Summary plots of the influence of sexual reproduction on the evolution of enset. (A) co-ancestry heatmap based on shared haplotypes produced where each row corresponds to a 'recipient' and each column to a 'donor', (B) heatmap of the percentage of significant D-statistics grouped by the sample types occurring at positions P2 and P3 and (C) PCA with observed data and simulated hybrids between domesticated (DxD), domesticated and wild (DxW) and backcrosses between domesticated × wild hybrids and domesticated (D1xW_BC_D) after filtering sites with a minor allele frequency of < 0.05.

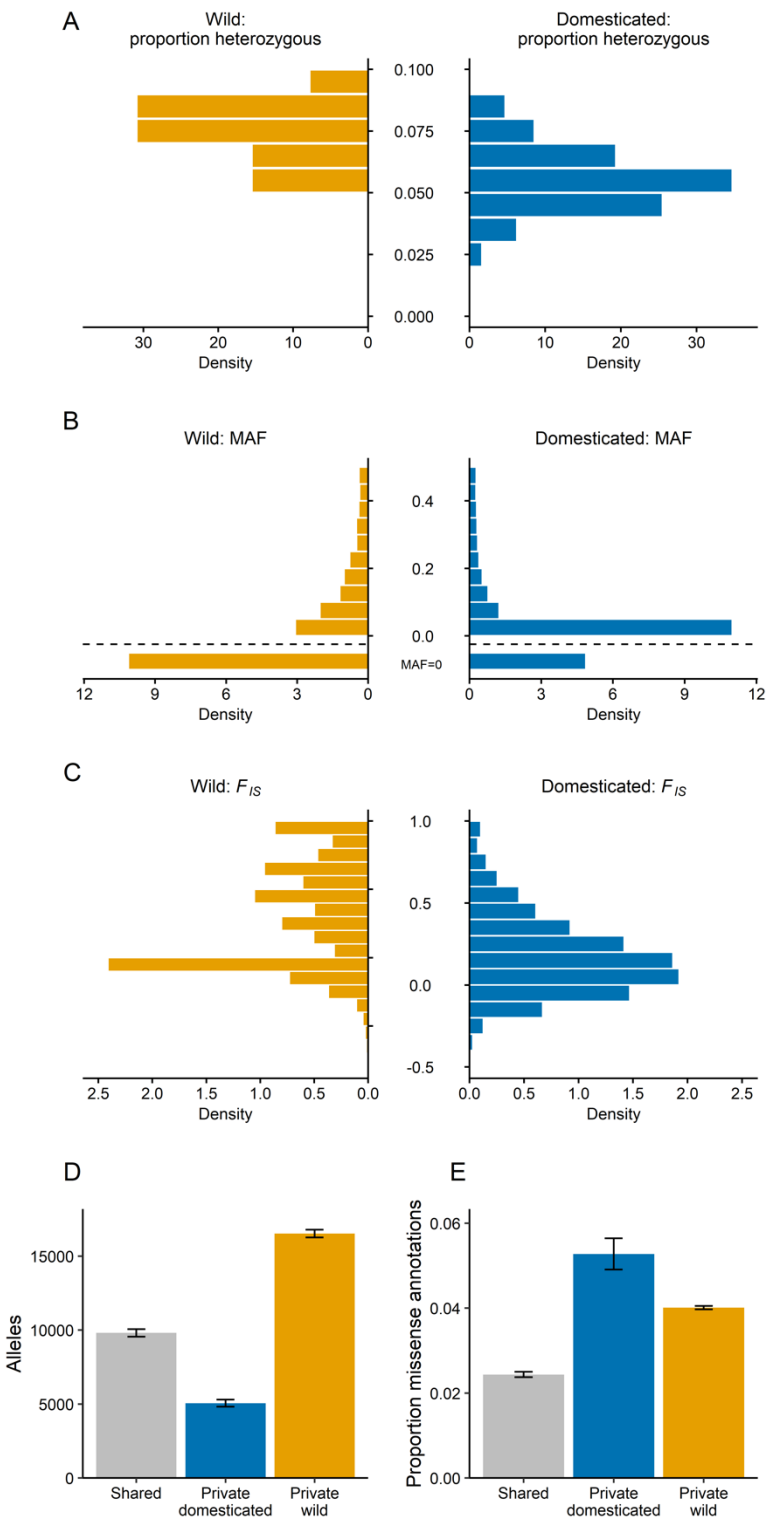

Figure 5 Population genetic analyses of wild and domesticated lineages. (A) proportion of heterozygous sites per individual, (B) minor allele frequency (MAF) spectrum with a separate bar for a MAF of zero and (C) inbreeding coefficient (F_{IS}) after removing low frequency sites (MAF < 0.1) for the wild and domesticated populations respectively. (D) Number of alleles identified as of shared, private domesticated and private wild and (E) proportion of missense annotations per allele category, calculated using rarefaction with error bars for standard deviation.

Tables

Table 1 Tests for phylogenetic signal for landrace traits as well as random and BM modelled traits using Abouheif's C_{mean} and Pagel's λ. P values less than 0.05 are shown in bold.

	Cmean	p	Pagel's λ	p
Bulla	0.117	0.009	0.154	0.055
Kocho	0.120	0.007	0.176	0.006
Fibre	0.150	**0.003**	0.205	**0.001**
Medicine	0.049	0.135	0.105	0.221
EBW susceptible	0.125	0.005	0.190	**0.001**
EBW tolerant	0.230	**0.001**	0.235	**0.001**
Random	-0.062	0.876	0.000	1.000
BM	0.815	**0.001**	1.008	**0.001**

Supporting Information

Methods

Figures

Tables

References

Alexa, A., & Rahnenfuhrer, J. (2021). *topGO: Enrichment Analysis for Gene Ontology*.

Balloux, F., Lehmann, L., & De Meeûs, T. (2003). The population genetics of clonal and partially clonal diploids. *Genetics*, *164*(4), 1635–1644. doi: 10.2135/cropsci1967.0011183X000700040005x

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*(1), 289–300.

Blomme, G., Yemataw, Z., Tawle, K., Sinohin, V., Gueco, L., Kebede, R., & Lalusin, A. (2018). *Assessing enset fibre yield and quality for a wide range of enset [Ensete ventricosum (Welw .) Cheesman] landraces in Ethiopia*. *73*(6), 328–341.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. doi: 10.1093/bioinformatics/btu170

Borrell, J. S., Biswas, M. K., Goodwin, M., Blomme, G., Schwarzacher, T., Heslop-Harrison, P. J. S., … Wilkin, P. (2019). Enset in Ethiopia: a poorly characterised but resilient starch staple. *Annals of Botany*, *123 (5)*, 747–766. doi: 10.1093/aob/mcy214

Borrell, J. S., Gebremariam, Z., & Wendawek, A. M. (2021). Utilize existing genetic diversity before genetic modification in indigenous crops. *Nature Biotechnology*, *39*(9), 1064–1065. doi: https://doi.org/10.1038/s41587-021-01048-6

Borrell, J. S., Goodwin, M., Blomme, G., Jacobsen, K., Wendawek, A. M., Gashu, D., … Gardens, R. B. (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*, (August 2019), 1–17. doi: 10.1002/ppp3.10084

Brandt, S., Spring, A., Hiebsch, C., McCabe, J. T., Tabogie, E., Wolde-Michael, G., … Tesfaye, S. (1997).

The “Tree Against Hunger”: Enset-Based Agricultural Systems in Ethiopia. In *Advancement Of Science*. doi: American Association for the Advancement of Science

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. doi: 10.1111/mec.12354

Charlesworth, B., Betancourt, A. J., Kaiser, V. B., & Gordo, I. (2009). Genetic recombination and molecular evolution. *Cold Spring Harbor Symposia on Quantitative Biology*, *74*(Robertson 1961), 177–186. doi: 10.1101/sqb.2009.74.015

Chen, L. Y., VanBuren, R., Paris, M., Zhou, H., Zhang, X., Wai, C. M., … Ming, R. (2019). The bracteatus pineapple genome and domestication of clonally propagated crops. *Nature Genetics*, *51*(10), 1549–1558. doi: 10.1038/s41588-019-0506-8

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., … Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92. doi: 10.4161/fly.19695

Comeron, J. M., Williford, A., & Kliman, R. M. (2008). The Hill-Robertson effect: Evolutionary consequences of weak selection and linkage in finite populations. *Heredity*, *100*(1), 19–31. doi: 10.1038/sj.hdy.6801059

Darriba, D., Posada, D., Flouri, T., Kozlov, A. M., Stamatakis, A., & Morel, B. (2019). *ModelTest-NG : A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models*. *37*(1), 291–294. doi: 10.1093/molbev/msz189

Denham, T., Barton, H., Castillo, C., Crowther, A., Dotte-Sarout, E., Florin, S. A., … Fuller, D. Q. (2020). The domestication syndrome in vegetatively propagated field crops. *Annals of Botany*, *125*(4),

581–597. doi: 10.1093/aob/mcz212

Dodd, R. S., & Douhovnikoff, V. (2016). Adjusting to Global change through clonal growth and epigenetic variation. *Frontiers in Ecology and Evolution*, *4*(JUL), 1–6. doi: 10.3389/fevo.2016.00086

Dorant, Y., Cayuela, H., Wellband, K., Laporte, M., Rougemont, Q., Mérot, C., … Bernatchez, L. (2020). Copy number variants outperform SNPs to reveal genotype–temperature association in a marine species. *Molecular Ecology*, *29*(24), 4765–4782. doi: 10.1111/mec.15565

Doyle, J., Doyle, J. L. (1987). Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem Bull*, *19*(11), 11–15.

Droc, G., Larivière, D., Guignon, V., Yahiaoui, N., This, D., Garsmeur, O., … Bocs, S. (2013). The banana genome hub. *Database*, *2013*, 1–14. doi: 10.1093/database/bat035

Eshetae, M. A., Hailu, B. T., & Demissew, S. (2021). Spatial characterization and distribution modelling of Ensete ventricosum (wild and cultivated) in Ethiopia. *Geocarto International*, *36*(1), 60–75. doi: 10.1080/10106049.2019.1588392

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, *180*(2), 977–993. doi: 10.1534/genetics.108.092221

Foster, T. M., & Aranzana, M. J. (2018). Attention sports fans! the far-reaching contributions of bud sport mutants to horticulture and plant biology. *Horticulture Research*, *5*(1). doi: 10.1038/s41438-018-0062-x

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

Haile, B., & Tesfaye, B. (2022). Indigenous Biosystematics of Enset (Ensete ventricosum [Welw.] Cheesman) in its Center of Origin and Diversity, Southwest Ethiopia: Folk Nomenclature, Classification, and Descriptors. In *Economic Botany*. Springer US. doi: 10.1007/s12231-022-09553-8

Halkett, F., Simon, J. C., & Balloux, F. (2005). Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology and Evolution*, *20*(4), 194–201. doi: 10.1016/j.tree.2005.01.001

Hodson, D. P., Jaleta, M., Tesfaye, K., Yirga, C., Beyene, H., Kilian, A., … Erenstein, O. (2020). Ethiopia's transforming wheat landscape: tracking variety use through DNA fingerprinting. *Scientific Reports*, *10*(1), 1–13. doi: 10.1038/s41598-020-75181-8

Hunduma, T., Kassahun, S., Hilu, E., & Oli, M. (2015). Evaluation of Enset Clones Resistance against Enset Bacterial Wilt Disease (Xanthomonas campestris pv. musacearum). *Journal of Veterinary Science & Technology*, *06*(03). doi: 10.4172/2157-7579.1000232

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254–267. doi: 10.1093/molbev/msj030

International Plant Genetic Resources Instit. (1996). *Descriptors for Banana (Musa spp.)*. Rome. Retrieved from https://books.google.co.id/books?id=l3mFQgAACAAJ

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

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. doi: 10.1093/bioinformatics/btr521

Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). *Novel R tools for analysis of genome-wide*

population genetic data with emphasis on clonality. *6*(June), 1–10. doi: 10.3389/fgene.2015.00208

Karamura, D. A., Karamura, E., & Tinzaara, W. (2012). *Banana cultivar names, synonyms and their usage in Eastern Africa*. Uganda: Bioversity International.

Keck, F., Rimet, F., Bouchez, A., & Franc, A. (2016). Phylosignal: An R package to measure, test, and explore the phylogenetic signal. *Ecology and Evolution*, *6*(9), 2774–2780. doi: 10.1002/ece3.2051

Kidane, S. A., Meressa, B. H., Haukeland, S., Hvoslef-Eide, A. K. (Trine), & Coyne, D. L. (2021). The Ethiopian staple food crop enset (Ensete ventricosum) assessed for the first time for resistance against the root-lesion nematode Pratylenchus goodeyi. *Nematology*, *0*, 1–9. doi: 10.1163/15685411-bja10075

Koch, O., Mengesha, W. A., Pironon, S., Pagella, T., Ondo, I., Rosa, I., … Borrell, J. S. (2022). Modelling potential range expansion of an underutilised food security crop in Sub-Saharan Africa. *Environmental Research Letters*, *17*(1). doi: 10.1088/1748-9326/ac40b2

Kozlov, A., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2018). RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. *BioRxiv*, 447110. doi: 10.1101/447110

Le Cam, S., Daguin-Thiébaut, C., Bouchemousse, S., Engelen, A. H., Mieszkowska, N., & Viard, F. (2019). A genome-wide investigation of the worldwide invader Sargassum muticum shows high success albeit (almost) no genetic diversity. *Evolutionary Applications*, (April 2019), 500–514. doi: 10.1111/eva.12837

Lemoine, F., Domelevo Entfellner, J. B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., & Gascuel, O. (2018). Renewing Felsenstein's phylogenetic bootstrap in the era of Big Data.

Nature, *556*(7702), 452–456. doi: 10.1002/jmri.25711.PET/MRI

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*(5), 589–595. doi: 10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. doi: 10.1093/bioinformatics/btp352

Lutteropp, S., Kozlov, A. M., & Stamatakis, A. (2019). A fast and memory-efficient implementation of the transfer bootstrap. *Bioinformatics*, (November), 1–2. doi: 10.1093/bioinformatics/btz874

Malinsky, M., Matschiner, M., & Svardal, H. (2020). Dsuite - fast D -statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, 0–2. doi: 10.1111/1755-0998.13265

Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018). RADpainter and fineRADstructure: Population Inference from RADseq Data. *Molecular Biology and Evolution*, *35*(5), 1284–1290. doi: 10.1093/molbev/msy023

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, *17.1*, 10–12.

Mckey, D., Elias, M., Pujol, M. E., & Duputié, A. (2010). The evolutionary ecology of clonally propagated domesticated plants. *New Phytologist*, *186*(2), 318–332. doi: 10.1111/j.1469-8137.2010.03210.x

McSweeney, C., New, M., & Lizcano, G. (2010). *UNDP Climate Change Country Profiles*.

Meyer, R. S., Duval, A. E., & Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytologist*, *196*, 29–48. doi: 10.1111/j.1469-8137.2012.04253.x

Miller, A. J., & Gross, B. L. (2011). From forest to field: Perennial fruit crop domestication. *American Journal of Botany*, *98*(9), 1389–1414. doi: 10.3732/ajb.1000522

Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., & Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, *3*(4), 743–756. doi: 10.1111/j.2041-210X.2012.00196.x

Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., … Buckler, E. S. (2011). Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 3530–3535. doi: 10.1073/pnas.1009363108

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. doi: 10.1093/molbev/msu300

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., … Wagner, H. (2019). *Package "vegan": Community ecology package* (p. http://vegan.r-forge.r-project.org/). p. http://vegan.r-forge.r-project.org/.

Ott, A., Liu, S., Schnable, J. C., Yeh, C. T., Wang, K. S., & Schnable, P. S. (2017). tGBS® genotyping-by-sequencing enables reliable genotyping of heterozygous loci. *Nucleic Acids Research*, *45*(21). doi: 10.1093/nar/gkx853

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. doi: 10.1093/bioinformatics/btq033

R Core team. (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rabbi, I. Y., Kulakow, P. A., Manu-Aduening, J. A., Dankyi, A. A., Asibuo, J. Y., Parkes, E. Y., … Maredia,

M. K. (2015). Tracking crop varieties using genotyping-by-sequencing markers: A case study using cassava (Manihot esculenta Crantz). *BMC Genetics*, *16*(1), 1–11. doi: 10.1186/s12863-015-0273-1

Ramu, P., Esuma, W., Kawuki, R., Rabbi, I. Y., Egesi, C., Bredeson, J. V., … Lu, F. (2017). Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. *Nature Genetics*, *49*(6), 959–963. doi: 10.1038/ng.3845

Reichel, K., Masson, J. P., Malrieu, F., Arnaud-Haond, S., & Stoeckel, S. (2016). Rare sex or out of reach equilibrium? The dynamics of FIS in partially clonal organisms. *BMC Genetics*, *17*(1), 1–16. doi: 10.1186/s12863-016-0388-z

Ruas, M., Guignon, V., Sempere, G., Sardos, J., Hueber, Y., Duvergey, H., … Rouard, M. (2017). MGIS: managing banana (Musa spp.) genetic resources information and high-throughput genotyping data. *Database : The Journal of Biological Databases and Curation*, *2017*(May), 1–12. doi: 10.1093/database/bax046

Sardos, J., Mckey, D., Duval, M. F., Malapa, R., Noyer, J. L., & Lebot, V. (2008). Evolution of cassava (Manihot esculenta Crantz) after recent introduction into a South Pacific Island system: the contribution of sex to the diversification of a clonally propagated crop. *Genome*, *51*, 912–921. doi: 10.1139/G08-080

Sardos, J., Perrier, X., & Dole, J. (2016). DArT whole genome profiling provides insights on the evolution and taxonomy of edible Banana (Musa spp .) DArT whole genome profiling provides insights on the evolution and taxonomy of edible Banana (Musa spp .). *Annals of Bot*, *118*(7), 1269–1278. doi: 10.1093/aob/mcw170

Scarcelli, N., Tostain, S., Vigouroux, Y., Agbangla, C., Daïnou, O., & Pham, J. L. (2006). Farmers' use of wild relative and sexual reproduction in a vegetatively propagated crop. The case of yam in

Benin. *Molecular Ecology*, *15*(9), 2421–2431. doi: 10.1111/j.1365-294X.2006.02958.x

Tamrat, S., Borrell, J. S., Biswas, M. K., Gashu, D., Wondimu, T., & Carlos, A. (2020). Micronutrient composition and microbial community analysis across diverse landraces of the Ethiopian orphan crop enset. *Food Research International*, *137*.

Tamrat, S., Borrell, J. S., Shiferaw, E., Wondimu, T., Kallow, S., Davies, R. M., … Wilkin, P. (2022). Reproductive biology of wild and domesticated Ensete ventricosum: Further evidence for maintenance of sexual reproductive capacity in a vegetatively propagated perennial crop. *Plant Biology*, *24*(3), 482–491. doi: 10.1111/plb.13390

Tesfamicael, K. G., Gebre, E., March, T., Sznajder, B., Mather, D., & Marcelino Rodriguez Lopez, C. (2020). Accumulation of mutations in genes associated with sexual reproduction contributed to the domestication of a vegetatively propagated staple crop, enset. *Horticulture Research*, *7*(185), 1–31. doi: 10.1101/2020.04.01.020750

Wang, J., Fernández-Pavía, S. P., Larsen, M. M., Garay-Serrano, E., Gregorio-Cipriano, R., Rodríguez-Alvarado, G., … Goss, E. M. (2017). High levels of diversity and population structure in the potato late blight pathogen at the Mexico centre of origin. *Molecular Ecology*, *26*(4), 1091–1107. doi: 10.1111/mec.14000

Wu, D., Lao, S., & Fan, L. (2021). De-Domestication : An Extension of Crop Evolution. *Trends in Plant Science*. doi: 10.1016/j.tplants.2021.02.003

Yemataw, Z., Chala, A., Ambachew, D., Studholme, D., Grant, M., & Tesfaye, K. (2017). Morphological variation and inter-relationships of quantitative traits in enset (Ensete ventricosum (welw.) Cheesman) germplasm from South and South-Western Ethiopia. *Plants*, *6*(4), 56. doi: 10.3390/plants6040056

Yemataw, Z., Muzemil, S., Ambachew, D., Tripathi, L., Tesfaye, K., Chala, A., … Studholme, D. J.

(2018). Genome sequence data from 17 accessions of Ensete ventricosum, a staple food crop for millions in Ethiopia. *Data in Brief*, *18*, 285–293. doi: 10.1016/j.dib.2018.03.026

Yemataw, Z., Tesfaye, K., Zeberga, A., & Blomme, G. (2016). Exploiting indigenous knowledge of subsistence farmers' for the management and conservation of Enset (Ensete ventricosum (Welw.) Cheesman) (musaceae family) diversity on-farm. *Journal of Ethnobiology and Ethnomedicine*, *12*(1), 34. doi: 10.1186/s13002-016-0109-8

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. doi: 10.1111/2041-210X.12628

Zhou, Y., Massonnet, M., Sanjak, J. S., Cantu, D., & Gaut, B. S. (2017). Evolutionary genomics of grape (Vitis vinifera ssp. vinifera) domestication. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(44), 11715–11720. doi: 10.1073/pnas.1709257114

Zohary, D. (2004). Unconscious selection and the evolution of domesticated Plants. *Economic Botany*, *58*(1), 5–10.