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

White, Oliver W.; Biswas, Manosh Kumar; Abebe, Wendawek M.; Dussert, Yann; Kadebe, Firew ; Nichols, Richard A.; Buggs, Richard J.A.; Demissew, Sabsebe; Woldeyes, Feleke; Papadopulos, Alexander S. T.; Schwarzacher, Trude; Heslop-Harrison, Pat J.S.; Wilkin, Paul; Borrell, James S.

Molecular Ecology

DOI:
10.1111/mec.17033

E-pub ahead of print: 02/06/2023

Publisher's PDF, also known as Version of record

Received: 1 September 2022 | Revised: 24 April 2023 | Accepted: 15 May 2023

DOI: 10.1111/mec.17033

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

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

Oliver W. White[1] | Manosh Kumar Biswas[2] | Wendawek M. Abebe[3] | Yann Dussert[4,5] | Firew Kebede[3] | Richard A. Nichols[5] | Richard J. A. Buggs[1,5] | Sebsebe Demissew[6] | Feleke Woldeyes[7] | Alexander S. T. Papadopulos[8] | Trude Schwarzacher[2,9] | Pat J. S. Heslop-Harrison[2,9] | Paul Wilkin[1] | James S. Borrell[1]

[1]Royal Botanic Gardens Kew, Richmond, UK

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

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

[4]CNRS, Université de Poitiers, EBI, Poitiers, France

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

[6]Department of Biology, Addis Ababa University, Addis Ababa, Ethiopia

[7]Ethiopian Biodiversity Institute, Addis Ababa, Ethiopia

[8]Schools of Natural Sciences, University of Bangor, Gwynedd, UK

[9]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, China

Correspondence
James S. Borrell, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, UK.
Email: j.borrell@kew.org

Funding information
GCRF Agrisystems award, Grant/Award Number: BB/S014896/1; GCRF Foundation Awards for Global Agricultural and Food Systems Research, Grant/Award Number: BB/P02307X/1; GCRF I-FLIP grant, Grant/Award Number: BB/S018980/1

Handling Editor: Ana Caicedo

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 characterize the domestication and contemporary genetic diversity of Enset (*Ensete ventricosum*), an indigenous African relative of bananas (*Musa*) and a principal starch staple for 20 million Ethiopians. Wild enset reproduction occurs strictly by sexual outcrossing, but for cultivation, it is propagated clonally and associated with diversification and specialization 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, an excess of putatively deleterious missense mutations at low frequency present as heterozygotes suggested an 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 the amplification of favourable genotypes through an extended period of clonal propagation. Among domesticated lineages, we found a significant phylogenetic signal for multiple farmer-identified food, nutrition and disease resistance traits and little evidence of contemporary recombination. The 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, clonal evolution, domestication, *Ensete ventricosum*, Ethiopia, food security, genotyping-by-sequencing, tGBS

1 | INTRODUCTION

More than half of domesticated food crop species are clonally propagated (Meyer et al., 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 et al., 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 et al., 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 preventing or slowing 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). Enset's pseudostem and corm (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 generate lateral rhizome suckers (shoots arising from the roots underground), which occur in bananas. By contrast, domesticated enset is 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. Therefore, in contrast to many crops for which both reproductive methods are employed (McKey et al., 2010; Simmonds, 1997), wild and domesticated enset displays a long-term, consistent and analytically tractable distinction between sexual and clonal modes of evolution (Tamrat et al., 2022; Tesfamicael et al., 2020). Both wild and domesticated enset plants co-occur in situ in some areas, and thus, present a system in which the consequences of contrasting reproductive strategies can be investigated.

Defining a domestication syndrome in enset is challenging, as in many vegetative and perennial crops (Denham et al., 2020). The principal distinction is the palatability of the pseudostem and corm tissue, which is bitter and largely inedible in wild populations and sweet in domesticated populations. Other morphological distinctions include leaf, midrib, petiole and pseudostem colour (wild: green/glaucous, domesticated: green, red, yellow, purple, black), corm size (wild: small, domesticated: enlarged), corm colour (wild: dark, domesticated: cream to white) and the presence of wax on the ventral leaf blade (wild: absent, domesticated: often present) (Borrell et al., 2019). In addition to morphological differences, Birmeta et al. (2002) reported genetic differentiation between wild and domesticated plants. 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 phenotypically diverse, with >1200 named landraces occurring across a wider range of environments than any other local crop (Table S1) (Borrell et al., 2019; Eshetae et al., 2021; Friis et al., 2010). 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 et al., 2016), and pest (Kidane et al., 2021) and disease tolerance (Hunduma et al., 2015). Farmers recognize, value and utilize this diversity and often seek to maintain landraces with a range of trait values on their farm (Borrell et al., 2020; Olango et al., 2014).

Genetic variation in clonally propagated crops commonly originates from sexual populations before, during and after domestication (Myles et al., 2011). This includes the integration of volunteer seedlings resulting from recombination between domesticated landraces, as observed in cassava (Elias et al., 2000; Sardos et al., 2008), or the integration of seedlings resulting from sexual crosses between

wild and domesticated landraces, for example 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 et al., 2003) may become increasingly important (Foster & Aranzana, 2018; Zhou et al., 2017). We expect novel variation from these sources to accumulate over time and be detectable as increased heterozygosity in clonal enset lineages. While 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 domestication and diversification of enset. We first assess the differentiation of domesticated enset from partly sympatric wild populations and then characterize the diversity of agronomic traits across the enset phylogeny. Second, to improve our understanding of the enset domestication process, we used *D*-statistics to compare the frequency of sexual reproduction across wild and domesticated lineages. Finally, integrating across multiple analyses, we evaluate evidence for a past population bottleneck, as well as contemporary accumulation of novel putatively deleterious diversity in the absence of recombination. We use our findings to better understand this trade-off and the prospects for indigenous clonal agricultural systems under major future climate and food security challenges.

2 | METHODS

2.1 | Collection of enset landraces

We collected leaf samples of 225 domesticated enset individuals across the zone of enset cultivation in 2017–19 (Figure 1a). 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 accessions, which following farmers' perspectives, we termed semi-domesticated (described as 'naturalized' in Hildebrand, 2001). While semi-domesticated individuals occur outside of cultivation (and so might be considered 'wild'), they are identified by farmers as distinct from either domesticated enset or wild populations (Hildebrand, 2001), and as such farmers do not attribute them to a named landrace (Haile & Tesfaye, 2022). They are generally recorded near cultivated plots and may be used as a famine food, but otherwise their origins are uncertain. Finally, we included one horticultural accession (*Ensete ventricosum* 'Maurelii') and three outgroup samples comprising *E. livingstonianum*, closely related to the African *Ensete* species from the Living Collection at RBGKew, and *E. ventricosum* accessions from outside Ethiopia (Table S2).

2.2 | Characterizing enset trait diversity

During field collections, 1340 farmers were asked to report the landraces present on their farm that are preferred for important agronomic properties including quality of kocho (the main enset-derived food product), quality of bulla (an additional food product), medicinal value, fibre quality and enset bacterial wilt susceptibility or tolerance (Methods S3). The cumulative number of times a landrace was observed across all farms was also recorded. This enabled us to generate relative importance scores for landraces across a range of continuous traits. To minimize the potential for farmer or researcher misidentification, perhaps due to differences between ethnic groups and languages, trait values for samples were collated only from survey data collected within the same administrative zone and ethnic group as for a given genotyped landrace.

2.3 | Sequencing and variant calling

DNA was isolated from silica-dried leaf material using a modified CTAB-based protocol (Doyle & 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 et al., 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 et al., 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 Table S3. 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 (Methods S1). 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 recombination, whereas the second dataset, hereafter referred to as

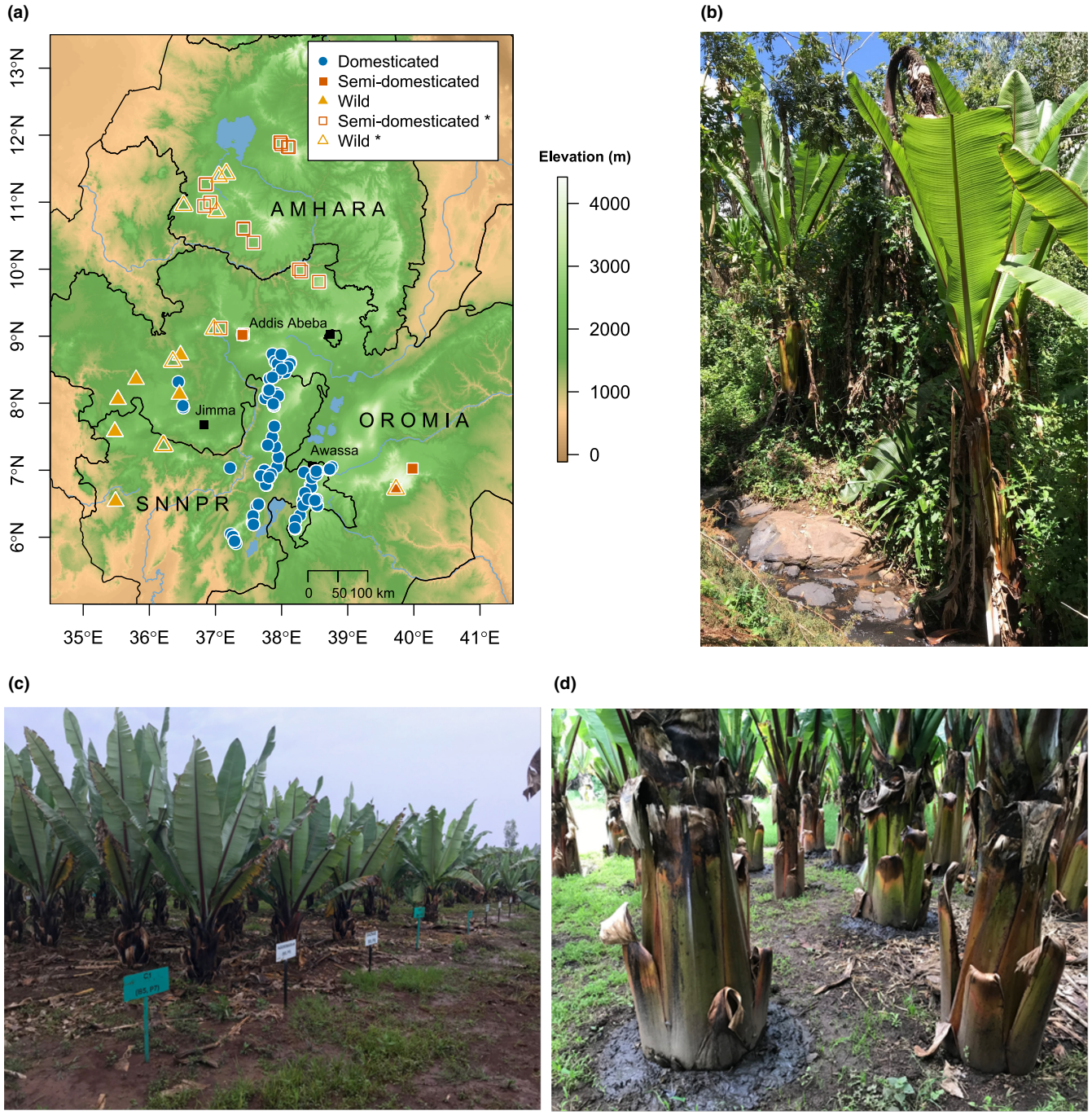

FIGURE 1 Distribution and phenotypic diversity of enset. (a) Distribution of domesticated, semi-domesticated and wild samples collected in Ethiopia for the present study. Points with no fill (marked with an asterisk in the legend) refer to further collection localities of wild and semi-domesticated samples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

the 'population genetic dataset', was used to investigate the evidence of a population bottleneck in domesticated enset.

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

For the *population genetic dataset*, we retained a single representative per multilocus genotype (MLG) for wild and domesticated samples (i.e. clone correction; see Methods: Identification of putative clones). Semi-domesticated and outgroup samples were excluded, resulting in a sample size of 141, comprising 128 domesticated and 13 wild samples. SNP calling was repeated as above, with the exception that SNPs were retained if present in 80% of all individuals, 80% of domesticated samples or 80% of wild samples. These filters were used to ensure that SNPs unique to either the wild or domesticated populations were

retained despite different sample sizes. SNPs were not filtered using a minimum allele count as we expect a fraction of somatic mutations to be present at very low frequency. In addition, we removed SNPs categorized as duplicated or diverged using the filtering procedures developed by Dorant et al. (2020). In general, we were cautious in our SNP filtering approach based on F_{IS}, which is expected to be negative under clonal variation due to an excess of heterozygotes relative to random mating.

2.4 | Phylogenetic analysis of enset diversity

To identify and visualize genetic clusters using the phylogenetic analysis dataset, we used principal component analysis (PCA) implemented using dudi.pca in the adegenet package (Jombart, 2008; Jombart & Ahmed, 2011; R Core Team, 2020). For phylogenetic reconstruction, we employed a maximum likelihood (ML) approach implemented in RAxML Next Generation v.0.9.0 (Kozlov et al., 2019) 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 5000 random and 5000 parsimony starting trees, selecting the tree with the lowest log-likelihood, before performing 5000 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 et al., 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 et al., 2015) (Methods S2), with topologies compared using phytools and visualized using ggtree (R Core Team, 2020; Yu et al., 2017), and a neighbour network using SplitsTree v.4.15.1 (Huson & Bryant, 2006).

2.5 | 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 et al., 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 be possible to define clonal individuals based solely on genetic identity due to somatic mutations and sequencing/SNP calling errors (Le Cam et al., 2020; Wang et al., 2017). We further identified MLGs with samples that were non-monophyletic on our RAXML-NG phylogenetic tree and removed these from downstream analyses. To estimate the total number of MLGs in the enset growing regions, we used an accumulation curve approach implemented in vegan (Oksanen et al., 2019). We randomly sampled genotypes across a range of sample sizes and calculated the number of MLGs. We then extrapolated using the best fitting model, estimating error over 100 replicate runs.

2.6 | Phylogenetic signal of landrace traits

After correction for the number of landrace observations, we tested for evidence of a phylogenetic signal in trait scores by calculating Abouheif's C_{mean} and Pagel's λ using the R package phylosignal (Keck et al., 2016). These indices differ in that they are based on principles of spatial autocorrelation and evolutionary models, respectively. However, both were found to be reliable measures of phylogenetic signal in simulation studies (Münkemüller et al., 2012). Null traits, modelled as random and under Brownian motion (BM) were added to our analysis to act as negative and positive controls, respectively.

2.7 | Frequency of sexual recombination

To identify the evidence of gene flow between lineages, we first visualized haplotype sharing patterns with RADpainter and fineRADstructure (Malinsky et al., 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 et al., 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 hybridized 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 positions and summarized 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

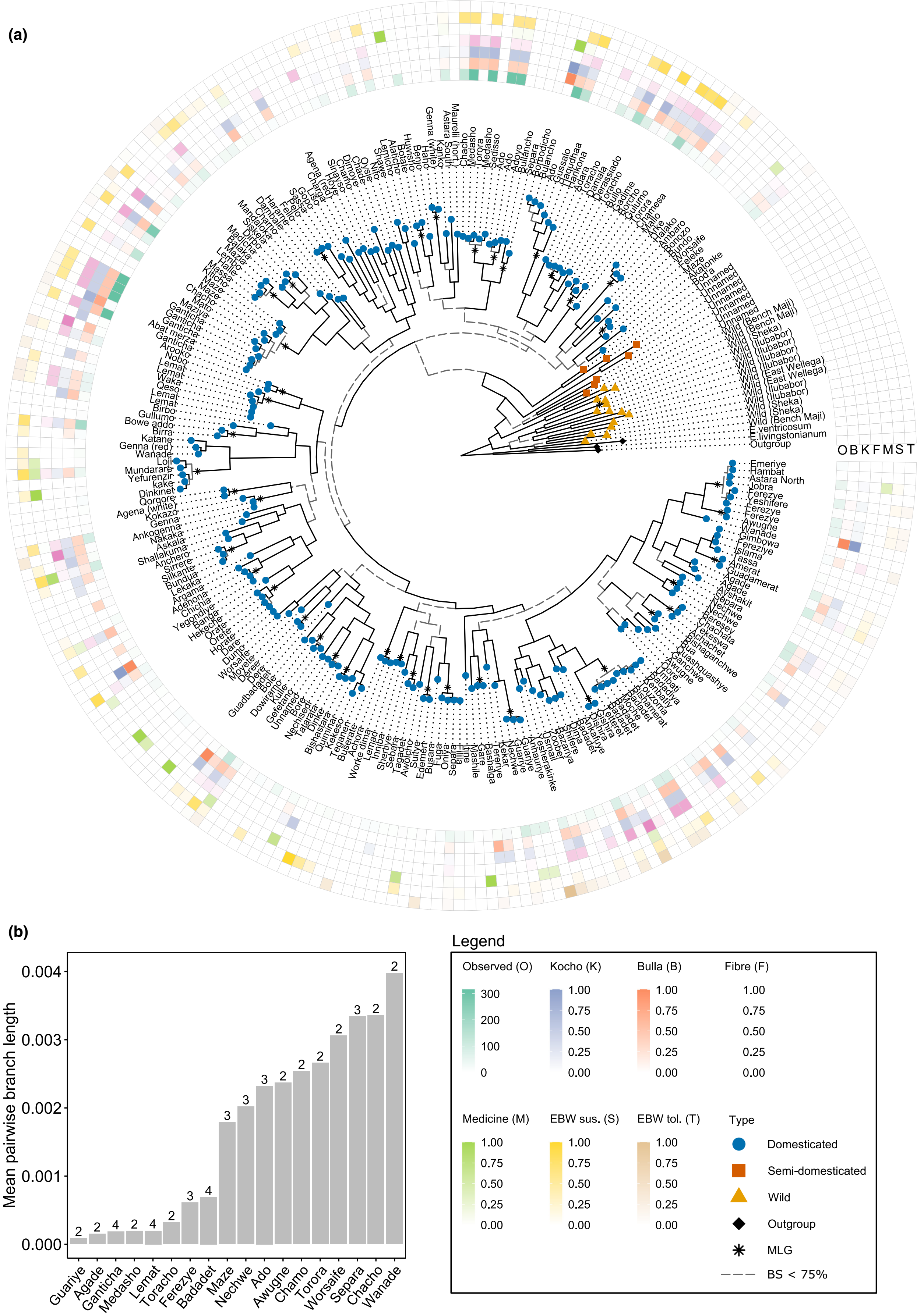

FIGURE 2 Phylogenetic and trait diversity of enset in Ethiopia. (a) Maximum likelihood phylogeny of enset produced using RAxML-NG with branch lengths shown. Branches with transfer bootstrap expectation (TBE) values <75% are shown as grey dashed lines. Branches leading to multilocus genotypes (MLGs) estimated 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.

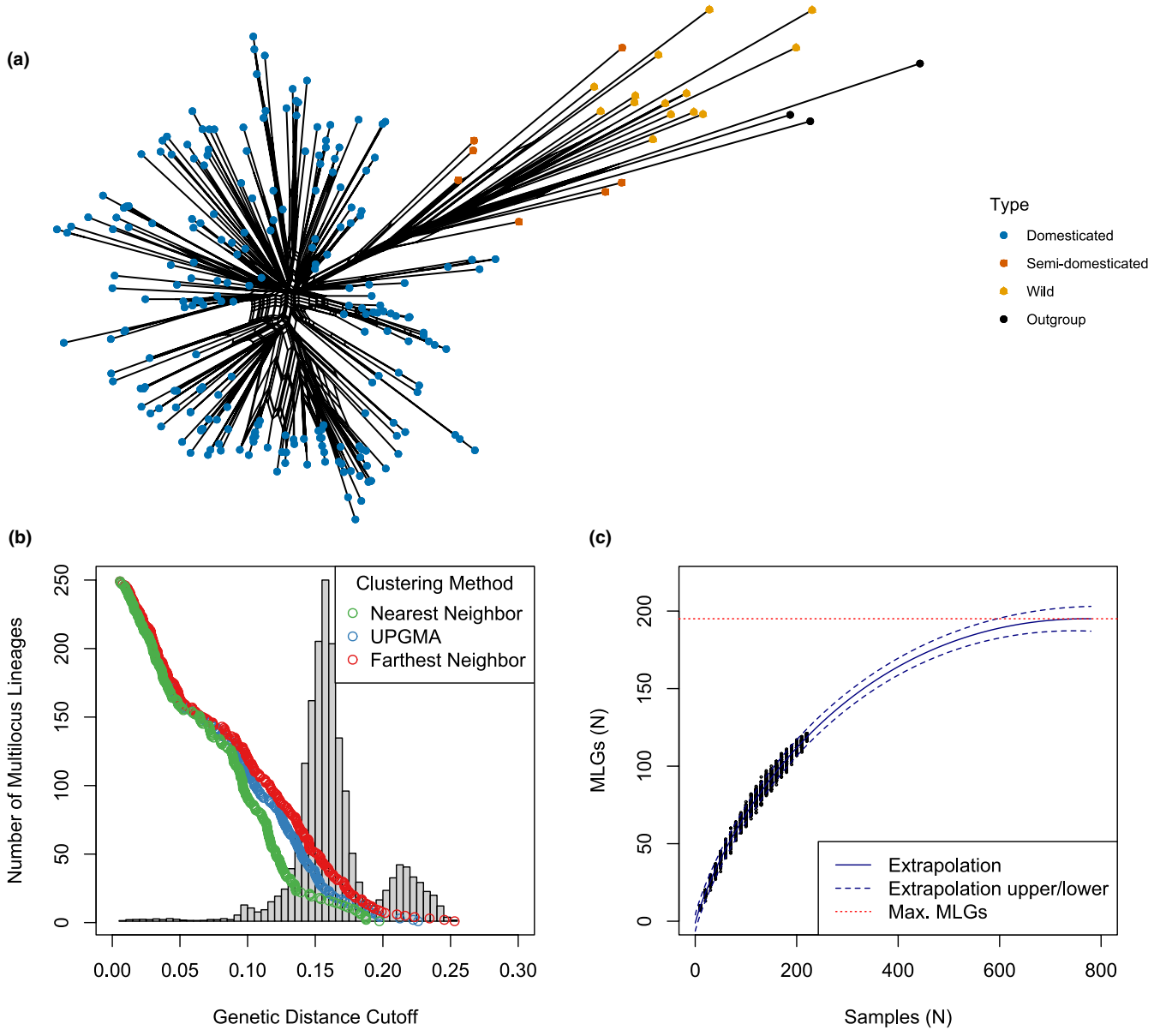

FIGURE 3 Genetic and multilocus genotype (MLG) diversity of enset. (a) Neighbour network of enset produced using SplitsTree with tips annotated by sample type. (b) Histogram of pairwise genetic distances. Thresholds for collapsing multilocus genotypes (MLGs) were: nearest 0.040, UPGMA 0.051 and farthest 0.051. 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.

F1 domesticated × wild crosses and domesticated. For the simulated crosses, a random subsample of 10 individuals was taken for the domesticated and wild populations, and 10 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.

2.8 | Evidence for a domestication bottleneck

Using the *population genetic dataset*, we calculated the proportion of heterozygous sites 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;

Enset traits	Cmean	*p*	Pagel's λ	*p*
Bulla quality	0.117	**.009**	0.154	.055
Kocho quality	0.120	**.007**	0.176	**.006**
Fibre quality	0.150	**.003**	0.205	**.001**
Medicinal value	0.049	.135	0.105	.221
EBW disease susceptibility	0.125	**.005**	0.190	**.001**
EBW disease tolerance	0.230	**.001**	0.235	**.001**
Random (negative control)	−0.062	.876	0.000	1.000
Brownian motion (positive control)	0.815	**.001**	1.008	**.001**

TABLE 1 Tests for phylogenetic signal for landrace traits as well as control (random and Brownian motion) traits using Abouheif's C_{mean} and Pagel's λ.

Note: Traits describe the famers' preferred landrace for each attribute. *p* Values <.05 are shown in bold.

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 were 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 were quantified using rarefaction as above.

3 | RESULTS

3.1 | Sequencing and variant calling

Sequencing generated approximately 787 M reads (Table S3), 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 1.67 M loci with a mean coverage of 15.6× per sample (range: 6.9–40.5). For the *phylogenetic dataset*, variant calling with populations resulted in 27,562 loci and 32,441 SNPs. After removing putative contaminant or paralogous sites, 9162 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 53,085 SNPs were retained after the removal of putative contaminant or paralogous sites.

3.2 | Phylogenetic analysis of enset diversity

Phylogenetic analyses with RAxML-NG (Figure 2a; Figure S1) found clear differentiation and strong support for monophyly of 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 (Figure S2) and neighbour networks (Figure 3a) recovered highly similar overall topologies and consistent domesticated subclades (Figure S3). 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 (Figure S4).

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 specialized for different purposes (Blomme et al., 2018; Borrell et al., 2020; Yemataw et al., 2017).

3.3 | 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 et al., 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 (Figure S5). A pair of wild

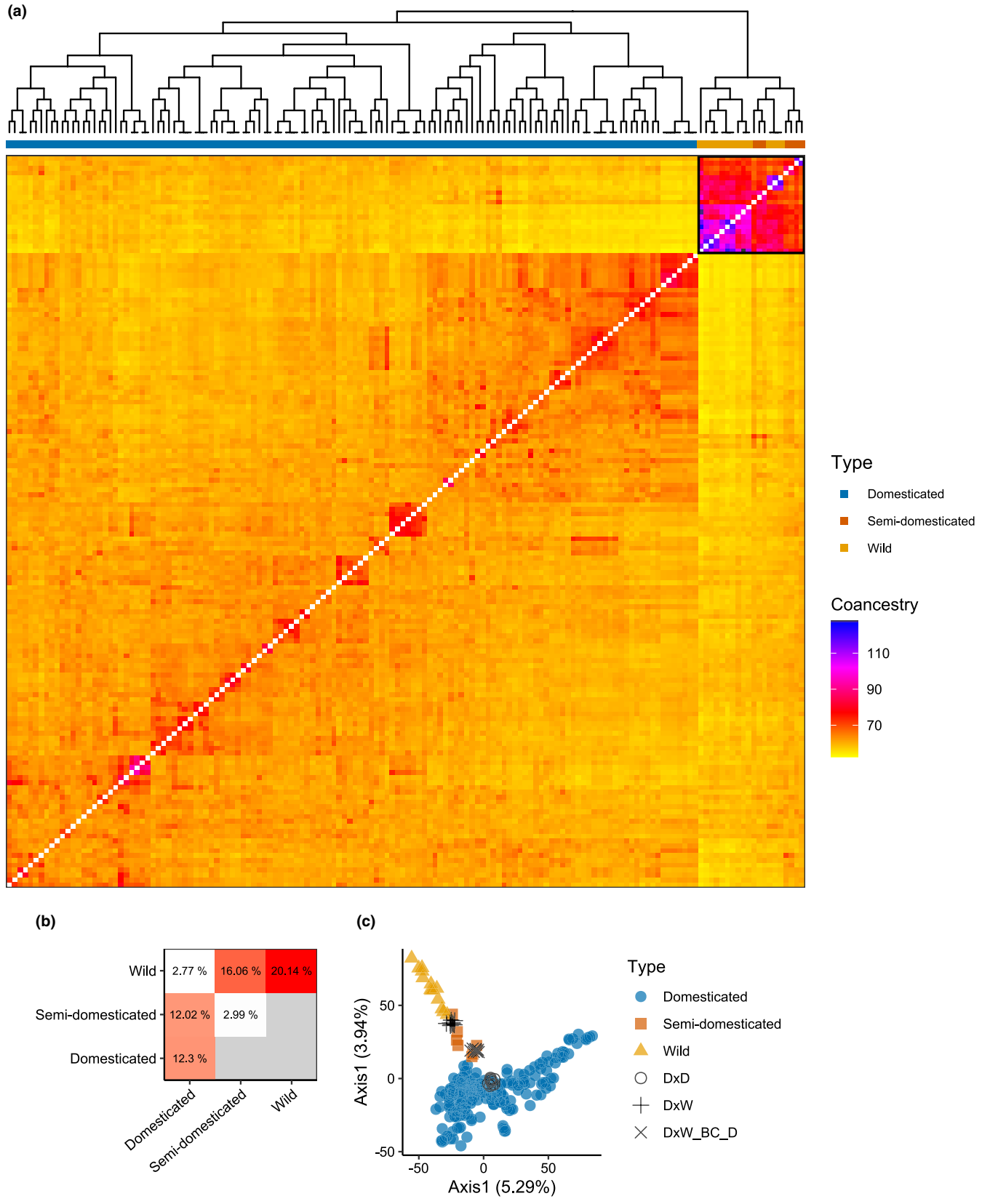

FIGURE 4 Summary plots of the influence of sexual reproduction on the evolution of enset. (a) co-ancestry heatmap based on shared haplotypes produced where each row corresponds to a 'recipient' and each column to a 'donor', (b) heatmap of the percentage of significant *D*-statistics grouped by the sample types occurring at positions P2 and P3 and (c) Principal component analysis with observed data and simulated hybrids between domesticated (D × D), domesticated and wild (DxW) and backcrosses between domesticated × wild hybrids and domesticated (D1 × W_BC_D) after filtering sites with a minor allele frequency of <0.05.

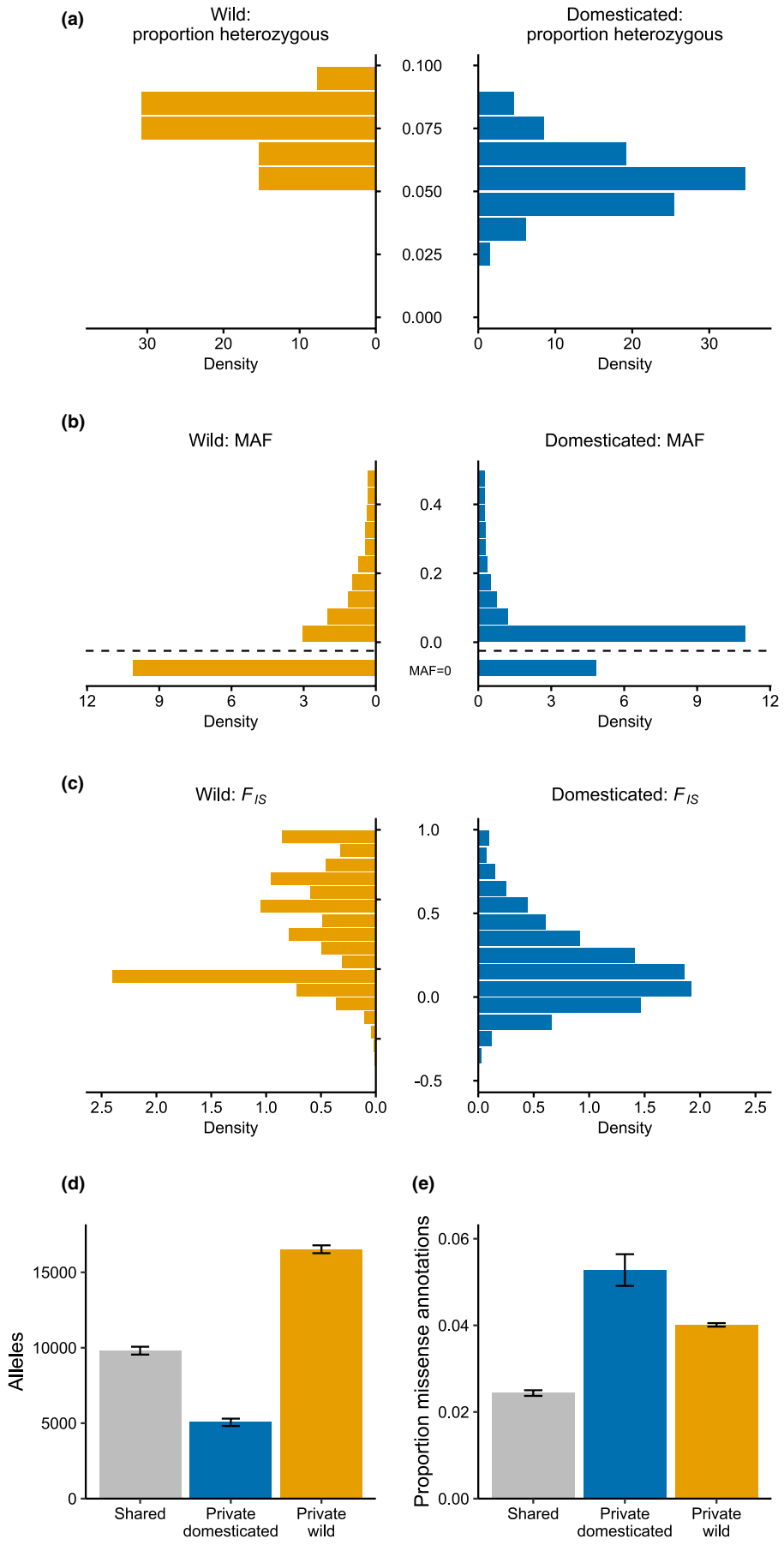

FIGURE 5 Population genetic analyses of wild and domesticated lineages. (a) proportion of heterozygous sites per individual, (b) minor allele frequency (MAF) spectrum with a separate bar for a MAF of zero and (c) inbreeding coefficient (F_{IS}) after removing low frequency sites (MAF < 0.1) for the wild 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.

samples were identified as putative clones by our analysis. However, it is unlikely that these wild samples were clones, rather siblings, and one was removed from all downstream analyses. 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).

3.4 | 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 hybridization ($D = 0.047$, $p < .01$). This topology was testing for the evidence of hybridization between wild (P3) and semi-domesticated (P2) or wild and domesticated (P1). Other population test topologies were not used as they were not possible based on the tree topology found by the phylogenetic analysis. Of 551,300 tests performed between individuals, 56,683 (10.28%) were significant after Benjamini and Hochberg correction for multiple tests (Table S4).

3.5 | Evidence for a domestication bottleneck

We found significantly lower observed heterozygosity in domesticated enset ($F_{(1,139)} = 16.55$, $p < .01$, 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 ($W = 1.76 \times 10^9$, $p < .01$, mean domesticated = 0.058, mean 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 estimate that approximately 37% of wild alleles were retained (Figure 5). We found that 16% of variation was unique to domesticated lineages, though we caution that our sampling of wild populations may be incomplete, and some populations may have become locally extinct due to extensive land-use change in Ethiopia.

We found significantly lower F_{IS} in domesticated lineages ($W = 2.42 \times 10^8$, $p < .01$, mean domesticated = 0.080, mean wild = 0.284; Figure 5c; Reichel et al., 2016). However, F_{IS} was not negative as expected for a clonal population. The lack of sexual recombination in clonal lineages results in non-random association between loci (Halkett et al., 2005). These associations make the selection against deleterious alleles less efficient due to effects known collectively as Hill-Robertson interference (Charlesworth et al., 2009; Comeron et al., 2008). Consistent with this process, we also find that alleles private to the domesticated populations have a significantly higher proportion of putatively deleterious missense mutations (Figure 5e), indicating that they are ineffectively purged from the genome and thus that deleterious alleles are accumulating in domesticated enset lineages. However, we note that these could be functionally important as part of the domestication process.

4 | DISCUSSION

Domestication has transformed enset from a wild, unpalatable, outcrossing giant forest herb, to the predominant staple starch crop for >20 million people. This has been achieved through the combination of sexual and clonal propagation, mediated by multiple stages of human selection. We hypothesise that the major domesticated lineages initially arose through historic sexual recombination. The existence of an early sexually recombining domesticated population ancestral to contemporary domesticated lineages is supported by the following evidence. First, we observe genetic differentiation between domesticated enset and wild populations. Direct farmer selection of clonal lineages with desirable attributes from wild ancestral populations (i.e. instantaneous domestication) is not supported by our phylogeny, because domesticated lines are not distributed among 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 subclades, 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 2–3 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). It is also plausible that the observed bottleneck is exacerbated by farmers' loss of clonal lineages, thus the diversity observed in the domesticated gene pool today is a subset of that which occurred in the initial phase of domestication, with farmers selecting among genotypes resulting from sexual recombination.

4.1 | 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 elevated low frequency 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 substitutions, and therefore that they are being less effectively purged. Over time, clonal propagation is therefore likely to introduce a limited amount of novel variation, as illustrated by variation within MLGs, but also an increased proportion of missense substitutions 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, which persists through the historic domestication bottleneck, though we identify 16% novel variation. This underlines the value of conserving wild populations, which might harbour agronomically useful genetic variation (Satori et al., 2022).

The rarity of hybridization between clonal lineages in domesticated enset is achieved through agronomic practices that 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 clonal propagation and introgression from wild populations (Wu et al., 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 localized to the periphery of the enset growing regions where germplasm management is substantially relaxed, enabling more frequent flowering and subsequent outcrossing. In addition, this hypothesis conforms to observations from Shigeta (1990), who documented crosses between cultivated and wild populations being incorporated into cultivated populations.

4.2 | Vernacular, genetic and trait diversity

We detected strong phylogenetic signal for enset traits (Figure 2). This suggests a predominantly genetic (rather than agronomic or environmental) basis to farmer-selected attributes and enables initial future characterization of poorly known landraces based on genotype (Borrell et al., 2021). In addition, evidence of a genetic basis for traits such as disease tolerance enable easier comparison of extensive regional Ethiopian literature that has assessed disease response in limited numbers of landraces (Gezahegn & Mekbib, 2016; Haile et al., 2020; Muzemil et al., 2021). This also highlights the challenge of trait analysis using only vernacular information. 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 et al., 2012), and discordance in landrace identification and genetic relationships also occurs in cassava (Rabbi et al., 2015) and Ethiopian wheat (Hodson et al., 2020). Consequently, our research highlights the need for a more formalized description of enset landrace diversity, as has been achieved for bananas (International Plant Genetic Resources Institute, 1996; Ruas et al., 2017), which in turn will aid experimental evaluation of landraces with agronomic traits of interest (e.g. Borrell et al., 2021; Hunduma et al., 2015; Kidane et al., 2021).

5 | 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. While 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 et al., 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.

AUTHOR CONTRIBUTIONS

James S. Borrell, Pat J. S. Heslop-Harrison, Sebsebe Demissew and Paul Wilkin designed the study. James S. Borrell, Wendawek M. Abebe and Firew Kebede performed the fieldwork collections. Pat J. S. Heslop-Harrison, Trude Schwarzacher and Manosh Kumar Biswas coordinated and led the DNA extraction and tGBS sequencing. Oliver W. White, Manosh Kumar Biswas and James S. Borrell led the bioinformatic analyses with input from Yann Dussert and Richard A. Nichols. All authors contributed to the interpretation of the results. Oliver W. White prepared the first draft and all authors contributed to the preparation of the final manuscript.

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 thank reviewers for their constructive comments on our initial 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.

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest.

FUNDING INFORMATION

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 AVAILABILITY STATEMENT

Raw sequence reads can be downloaded from the NCBI short read archive under BioProject PRJNA866136. Individual genotype data are available on DataDryad (https://doi.org/10.5061/dryad.3ffbg79ph). Associated sample metadata including georeferences are provided in Table S2.

BENEFIT-SHARING STATEMENT

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.

ORCID

Oliver W. White https://orcid.org/0000-0001-6444-0310
Manosh Kumar Biswas https://orcid.org/0000-0001-9228-1927
Yann Dussert https://orcid.org/0000-0003-1900-7715
Richard A. Nichols https://orcid.org/0000-0002-4801-9312
Richard J. A. Buggs https://orcid.org/0000-0003-4495-3738
Alexander S. T. Papadopulos https://orcid.org/0000-0001-6589-754X
Trude Schwarzacher https://orcid.org/0000-0001-8310-5489
Pat J. S. Heslop-Harrison https://orcid.org/0000-0002-3105-2167
Paul Wilkin https://orcid.org/0000-0003-4982-7175
James S. Borrell https://orcid.org/0000-0001-9902-7681

REFERENCES

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

Birmeta, G., Nybom, H., & Bekele, E. (2002). RAPD analysis of genetic diversity among clones of the Ethiopian crop plant *Ensete ventricosum*. *Euphytica*, *124*(3), 315–325. https://doi.org/10.1023/A:1015733723349

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. *Fruits*, *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. https://doi.org/10.1093/bioinformatics/btu170

Borrell, J. S., Biswas, M. K., Goodwin, M., Blomme, G., Schwarzacher, T., Heslop-Harrison, P. J. S., Wendawek, A. M., Berhanu, A., Kallow, S., Janssens, S., Molla, E. L., Davis, A. P., Woldeyes, F., Willis, K., Demissew, S., & Wilkin, P. (2019). Enset in Ethiopia: A poorly characterised but resilient starch staple. *Annals of Botany*, *123*(5), 747–766. https://doi.org/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. https://doi.org/10.1038/s41587-021-01048-6

Borrell, J. S., Goodwin, M., Blomme, G., Jacobsen, K., Wendawek, A. M., Gashu, D., Lulekal, E., Asfaw, Z., Demissew, S., Wilkin, P., & 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*, *2*, 212–228. https://doi.org/10.1002/ppp3.10084

Brandt, S., Spring, A., Hiebsch, C., McCabe, J. T., Tabogie, E., Wolde-Michael, G., Yntiso, G., Shigeta, M., & Tesfaye, S. (1997). *The "tree against hunger": Enset-based agricultural Systems in Ethiopia*. 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. https://doi.org/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*, 177–186. https://doi.org/10.1101/sqb.2009.74.015

Chen, L. Y., VanBuren, R., Paris, M., Zhou, H., Zhang, X., Wai, C. M., Yan, H., Chen, S., Alonge, M., Ramakrishnan, S., Liao, Z., Liu, J., Lin, J., Yue, J., Fatima, M., Lin, Z., Zhang, J., Huang, L., Wang, H., … Ming, R. (2019). The bracteatus pineapple genome and domestication of clonally propagated crops. *Nature Genetics*, *51*(10), 1549–1558. https://doi.org/10.1038/s41588-019-0506-8

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & 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. https://doi.org/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. https://doi.org/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. *Molecular Biology and Evolution*, *37*(1), 291–294. https://doi.org/10.1093/molbev/msz189

Denham, T., Barton, H., Castillo, C., Crowther, A., Dotte-Sarout, E., Florin, S. A., Pritchard, J., Barron, A., Zhang, Y., & Fuller, D. Q. (2020). The domestication syndrome in vegetatively propagated field crops. *Annals of Botany*, *125*(4), 581–597. https://doi.org/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*, 1–6. https://doi.org/10.3389/fevo.2016.00086

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

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*, 11–15.

Elias, M., Panaud, O., & Robert, T. (2000). Assessment of genetic variability in a traditional cassava (*Manihot esculenta* Crantz) farming system, using AFLP markers. *Heredity*, *85*, 219–230.

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. https://doi.org/10.1080/10106049.2019.1588392

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), 44. https://doi.org/10.1038/s41438-018-0062-x

Friis, I. B., Demissew, S., & van Breugel, P. (2010). *Atlas of the potential vegetation of Ethiopia* (Vol. *307*). Det Kongelige Danske Videnskabernes Selskab Copenhagen.

Gezahegn, G., & Mekbib, F. (2016). Evaluation of enset (*Ensete ventricosum* (Welw.) Cheesman) clone suckers to bacterial wilt disease pathogen under greenhouse condition. *Journal of Biology, Agriculture and Healthcare*, *6*(21), 51–56.

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

Haile, B., Fininsa, C., Terefe, H., Chala, A., & Hussen, S. (2020). Evaluation of enset (*Ensete ventricosum*) clones for resistance reaction against pathogenic *Xanthomonas campestris* pv. *Musacearum* isolates from southwestern Ethiopia. *Cogent Food and Agriculture*, *6*(1), 1773094. https://doi.org/10.1080/23311932.2020.1773094

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. *Economic Botany*, *76*, 382–413. https://doi.org/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. https://doi.org/10.1016/j.tree.2005.01.001

Hildebrand, E. H. (2001). Morphological characterization of domestic vs. forest-growing *Ensete ventricosum* (Welw.) Cheesman, Musaceae, in Sheko district, Bench-Maji Zone, Southwest Ethiopia. *Biologiske Skrifter*, *54*, 287–309.

Hodson, D. P., Jaleta, M., Tesfaye, K., Yirga, C., Beyene, H., Kilian, A., Carling, J., Disasa, T., Alemu, S. K., Daba, T., Alemayehu, Y., Badebo, A., Abeyo, B., & Erenstein, O. (2020). Ethiopia's transforming wheat landscape: Tracking variety use through DNA fingerprinting. *Scientific Reports*, *10*(1), 1–13. https://doi.org/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*, *6*(3), 1–7. https://doi.org/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. https://doi.org/10.1093/molbev/msj030

International Plant Genetic Resources Institute. (1996). *Descriptors for Banana (Musa spp.)*. International Plant Genetic Resources Institute.

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

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/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. *Frontiers in Genetics*, *6*(208), 1–10. https://doi.org/10.3389/fgene.2015.00208

Karamura, D. A., Karamura, E., & Tinzaara, W. (2012). *Banana cultivar names, synonyms and their usage in eastern Africa*. 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. https://doi.org/10.1002/ece3.2051

Kidane, S. A., Meressa, B. H., Haukeland, S., Hvoslef-Eide, A. K., & 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*, *23*(7), 771–779. https://doi.org/10.1163/15685411-bja10075

Koch, O., Mengesha, W. A., Pironon, S., Pagella, T., Ondo, I., Rosa, I., Wilkin, P., & Borrell, J. S. (2022). Modelling potential range expansion of an underutilised food security crop in sub-Saharan Africa. *Environmental Research Letters*, *17*(1), 14022. https://doi.org/10.1088/1748-9326/ac40b2

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A., & Wren, J. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*(21), 4453–4455. https://doi.org/10.1093/bioinformatics/btz305

Le Cam, S., Daguin-Thiébaut, C., Bouchemousse, S., Engelen, A. H., Mieszkowska, N., & Viard, F. (2020). A genome-wide investigation of the worldwide invader *Sargassum muticum* shows high success albeit (almost) no genetic diversity. *Evolutionary Applications*, *13*, 500–514. https://doi.org/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. https://doi.org/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. https://doi.org/10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lutteropp, S., Kozlov, A. M., & Stamatakis, A. (2019). A fast and memory-efficient implementation of the transfer bootstrap. *Bioinformatics*, *36*, 2280–2281. https://doi.org/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*, *21*, 584–595. https://doi.org/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. https://doi.org/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. https://doi.org/10.1111/j.1469-8137.2010.03210.x

McSweeney, C., New, M., & Lizcano, G. (2010). UNDP climate change country profiles: Ethiopia. Retrieved October 15, 2018, from https://www.geog.ox.ac.uk/research/climate/projects/undp-cp/

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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1111/j.2041-210X.2012.00196.x

Muzemil, S., Chala, A., Tesfaye, B., Studholm, D. J., Grant, M., Yemataw, Z., & Olango, T. (2021). Evaluation of 20 enset (*Ensete ventricosum*) landraces for response to *Xanthomonas vasicola* pv. *musacearum* infection. *European Journal of Plant Pathology*, *161*, 821–836. https://doi.org/10.1101/736793

Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., Prins, B., Reynolds, A., Chia, J.-M., Ware, D., Bustamante, C. D., & 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. https://doi.org/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. https://doi.org/10.1093/molbev/msu300

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

Olango, T. M., Tesfaye, B., Catellani, M., & Pè, M. E. (2014). Indigenous knowledge, use and on-farm management of enset (*Ensete ventricosum* (Welw.) Cheesman) diversity in Wolaita, Southern Ethiopia. *Journal of Ethnobiology and Ethnomedicine*, *10*(1), 41. https://doi.org/10.1186/1746-4269-10-41

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), e178. https://doi.org/10.1093/nar/gkx853

R Core team. (2020). *R: A language and environment for statistical computing (3.6.3)*. R Foundation for Statistical Computing.

Rabbi, I. Y., Kulakow, P. A., Manu-Aduening, J. A., Dankyi, A. A., Asibuo, J. Y., Parkes, E. Y., Abdoulaye, T., Girma, G., Gedil, M. A., Ramu, P., Reyes, B., & 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. https://doi.org/10.1186/s12863-015-0273-1

Ramu, P., Esuma, W., Kawuki, R., Rabbi, I. Y., Egesi, C., Bredeson, J. V., Bart, R. S., Verma, J., Buckler, E. S., & Lu, F. (2017). Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. *Nature Genetics*, *49*(6), 959–963. https://doi.org/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 F_{IS} in partially clonal organisms. *BMC Genetics*, *17*(1), 1–16. https://doi.org/10.1186/s12863-016-0388-z

Ruas, M., Guignon, V., Sempere, G., Sardos, J., Hueber, Y., Duvergey, H., Andrieu, A., Chase, R., Jenny, C., Hazekamp, T., Irish, B., Jelali, K., Adeka, J., Ayala-Silva, T., Chao, C. P., Daniells, J., Dowiya, B., Effa Effa, B., Gueco, L., … Rouard, M. (2017). MGIS: Managing banana (*Musa* spp.) genetic resources information and high-throughput genotyping data. *Database*, *2017*, 1–12. https://doi.org/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. https://doi.org/10.1139/G08-080

Sardos, J., Perrier, X., Doležel, J., Hřibová, E., Christelová, P., Van Den Houwe, I., Kilian, A., & Roux, N. (2016). DArT whole genome profiling provides insights on the evolution and taxonomy of edible banana (*Musa* spp.). *Annals of Botany*, *118*(7), 1269–1278. https://doi.org/10.1093/aob/mcw170

Satori, D., Tovar, C., Faruk, A., Hammond Hunt, E., Muller, G., Cockel, C., Kühn, N., Leitch, I. J., Lulekal, E., Pereira, L., Ryan, P., Willis, K. J., & Pironon, S. (2022). Prioritising crop wild relatives to enhance agricultural resilience in sub-Saharan Africa under climate change. *Plants People Planet*, *4*(3), 269–282. https://doi.org/10.1002/ppp3.10247

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. https://doi.org/10.1111/j.1365-294X.2006.02958.x

Shigeta, M. (1990). Folk In-situ conservation of *Ensete* [*Ensete ventricosum* (Welw.) E.E.Cheesman]: Toward the interpretation of indigenous Agricultural Science of the Ari, southwestern Ethiopia. *African Study Monographs*, *10*(3), 93–107.

Simmonds, N. W. (1997). A review of potato propagation by means of seed, as distinct from clonal propagation by tubers. *Potato Research*, *40*, 191–214.

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*, 109636.

Tamrat, S., Borrell, J. S., Shiferaw, E., Wondimu, T., Kallow, S., Davies, R. M., Dickie, J. B., Nuraga, G. W., White, O., Woldeyes, F., Demissew, S., & 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. https://doi.org/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. https://doi.org/10.1101/2020.04.01.020750

Wang, J., Fernández-Pavía, S. P., Larsen, M. M., Garay-Serrano, E., Gregorio-Cipriano, R., Rodríguez-Alvarado, G., Grünwald, N. J., & 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. https://doi.org/10.1111/mec.14000

Wu, D., Lao, S., & Fan, L. (2021). De-domestication: An extension of crop evolution. *Trends in Plant Science*, *26*(6), 560–574. https://doi.org/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. https://doi.org/10.3390/plants6040056

Yemataw, Z., Muzemil, S., Ambachew, D., Tripathi, L., Tesfaye, K., Chala, A., Farbos, A., O'Neill, P., Moore, K., Grant, M., & 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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1073/pnas.1709257114

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

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: White, O. W., Biswas, M. K., Abebe, W. M., Dussert, Y., Kebede, F., Nichols, R. A., Buggs, R. J. A., Demissew, S., Woldeyes, F., Papadopulos, A. S. T., Schwarzacher, T., Heslop-Harrison, P. J. S., Wilkin, P., & Borrell, J. S. (2023). Maintenance and expansion of genetic and trait variation following domestication in a clonal crop. *Molecular Ecology*, *00*, 1–16. https://doi.org/10.1111/mec.17033