G3, 2025, 15(2), jkae282
https://doi.org/10.1093/g3journal/jkae282
Advance Access Publication Date: 2 December 2024
Plant Genetics and Genomics

Evolutionary signatures of the erosion of sexual reproduction genes in domesticated cassava (*Manihot esculenta*)

Evan M. Long [1,2,*] Michelle C. Stitzer,[3] Brandon Monier [3] Aimee J. Schulz,[1] Maria Cinta Romay [3] Kelly R. Robbins [1] Edward S. Buckler [1,3,4]

[1]Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853, USA
[2]Department of Plant Sciences, University of California Davis, Davis, CA 95616, USA
[3]Institute for Genomic Diversity, Cornell University, Ithaca, NY 14853, USA
[4]United States Department of Agriculture-Agricultural Research Service, Robert W. Holley, Center for Agriculture and Health, Ithaca, NY 14853, USA

*Corresponding author: United States Department of Agriculture-Agricultural Research Service, N.W. Irrigation and Soils Research Lab, 3793 N. 3600 E. Kimberly, Idaho 83341-5076. Email: evan.long@usda.gov

Centuries of clonal propagation in cassava (*Manihot esculenta*) have reduced sexual recombination, leading to the accumulation of deleterious mutations. This has resulted in both inbreeding depression affecting yield and a significant decrease in reproductive performance, creating hurdles for contemporary breeding programs. Cassava is a member of the Euphorbiaceae family, including notable species such as rubber tree (*Hevea brasiliensis*) and poinsettia (*Euphorbia pulcherrima*). Expanding upon preliminary draft genomes, we annotated 7 long-read genome assemblies and aligned a total of 52 genomes, to analyze selection across the genome and the phylogeny. Through this comparative genomic approach, we identified 48 genes under relaxed selection in cassava. Notably, we discovered an overrepresentation of floral expressed genes, especially focused at 6 pollen-related genes. Our results indicate that domestication and a transition to clonal propagation have reduced selection pressures on sexually reproductive functions in cassava leading to an accumulation of mutations in pollen-related genes. This relaxed selection and the genome-wide deleterious mutations responsible for inbreeding depression are potential targets for improving cassava breeding, where the generation of new varieties relies on recombining favorable alleles through sexual reproduction.

Keywords: cassava; clonal reproduction; sexual reproduction; deleterious mutations; evolution; Plant Genetics and Genomics

Introduction

Cassava (*Manihot esculenta*) is a monoecious root crop grown in tropical regions around the world. Today, cassava is a major caloric source for over 500 million people, with a large number concentrated in sub-Saharan Africa (Parmar *et al.* 2017; Ferguson *et al.* 2019). Cassava is a woody shrub that naturally reproduces through outcrossing facilitated by separate male and female flowers. Although it is naturally perennial, it has been grown as an annual since its domestication 5–10 thousand years ago and vegetatively propagated through stem cuttings (Wang *et al.* 2014; Parmar *et al.* 2017). Centuries of selection have generated modern cassava varieties that produce large and abundant roots. This is particularly beneficial in sub-Saharan Africa where it is valued for its ability to grow with minimal inputs in marginally fertile lands, achieving an average of ~10 tons/hectare fresh root yield (Parmar *et al.* 2017). With continually rising demands from growing populations and impending difficulties due to climate change and other environmental considerations, breeding efforts for crop improvement in cassava have garnered increasing attention.

One hurdle that impedes contemporary breeding efforts in cassava is high levels of genetic load, made visible through heavy inbreeding depression and low reproductive fitness. Several studies quantified the level of inbreeding depression in self-fertilized cassava, such that a single generation of inbreeding can decrease fresh root yield by >60% (Rojas *et al.* 2009; de Freitas *et al.* 2016). These studies likely underestimate the impact of inbreeding depression, as only measured plants that successfully grew from self-fertilized seed can be measured, missing impacts on seed germination and sexual reproduction traits. Further evidence for poor sexual reproductive ability in cassava comes from high variability in flowering time, low numbers of female flowers, high rates of flower abortion, and low seed set (Ceballos *et al.* 2004; Souza *et al.* 2020; Oluwasanya *et al.* 2021). These limitations in seed production and viability limit breeders' abilities to make the successful crosses inherently necessary for developing new varieties and making genetic gains. Understanding genetic load can help researchers and breeders address these problems hindering the genetic improvement of cassava.

Genetic load, including that giving rise to inbreeding depression, can be measured through the accumulation of deleterious mutations throughout the genome. Cultivated cassava clones have thousands of deleterious mutations that are disproportionately maintained in a heterozygous state through clonal

Received on 18 September 2024; accepted on 18 November 2024

propagation, masking putatively recessive effects (Ramu *et al.* 2017; Long *et al.* 2023). Domestication can increase the fixation of deleterious mutations through linkage with selected loci, by "hitchhiking", but also by reducing the effective population size through population bottlenecks, thereby weakening the ability of selection to purge negative alleles (Moyers *et al.* 2018; Bosse *et al.* 2019). However, the genomic effects of domestication may differ in clonally propagated crops. Sexually reproducing annual crops can undergo thousands of generations of recombination as selection proceeds, but clonally propagated crops show fewer recombination events (Zhou *et al.* 2017; Chen, Huang, *et al.* 2019; Chen, VanBuren, *et al.* 2019). Thus, another factor that may contribute to the accumulation of deleterious mutations is the phenomenon known as "Muller's ratchet" (Muller 1964; Felsenstein 1974). Muller's ratchet explains the negative consequences of a lack of recombination, such as in clonal or asexual populations, where deleterious mutations are unable to be purged from the population's gene pool.

It is possible to evaluate deleterious mutations across the cassava genome, by placing it in a broader evolutionary context to detect conservation at these sites. Cassava belongs to the Euphorbiaceae, or spurge, family which is a very diverse clade within Malpighiales (Kubitzki 2014). There are over 8,000 species within the family ranging from tall trees like the rubber tree, *Hevea brasiliensis*, to the ornamental poinsettia, *Euphorbia pulcherrima*. Many species are acclimated to tropical regions; however, there are also species that are succulent and adapted to drier regions such as *Euphorbia canariensis*, or Canary Island Spurge. The few common features of uniovulate Euphorbiaceae species are "latex and laticifers, pollen morphology, and ovular and seed coat characters" (Kubitzki 2014). Cassava is known to have undergone a paleopolyploidy event that is shared with *H. brasiliensis* (Pootakham *et al.* 2017).

Although both the effect of domestication on deleterious mutations in plants and animals (Kantar *et al.* 2017; Kono *et al.* 2019), and the effects of clonal or asexual reproduction leading to the fixation of deleterious mutations have been extensively explored (Muller 1964; Dwivedi *et al.* 2023), the combination of the two in domesticated, asexually reproducing crops is rare. For example, although grapes and potatoes are clonally propagated, and maintain abundant deleterious mutations in a heterozygous state, the domestication bottleneck is not as strong as annual crops (Hardigan *et al.* 2017; Zhou *et al.* 2017). In contrast, cassava experienced a severe domestication bottleneck alongside a shift to clonal propagation that has been correlated to abundant deleterious mutations and severe inbreeding depression (Pujol *et al.* 2005; Ramu *et al.* 2017). While deleterious mutations are generally negatively correlated with the recombination rate, these mutations in cassava were not correlated, likely due to the lack of recombination from clonal propagation. In this study, we aim to use evolutionary conservation to measure the impact of deleterious mutations across the cassava genome.

Using 27 recently sequenced and assembled Euphorbiaceae species (Long *et al.* 2023), we perform genetic comparisons across a total of 52 species to detect selection signatures in cassava. This deep evolutionary comparison to species within the Euphorbiaceae family allows us to measure the impact of derived mutations on cassava genes and use intraspecific variation within cultivated cassava to detect how these mutations may be shaped by selection. In this study, we use these varied measures of evolutionary time to address the possible genomic consequences of clonal propagation across the cassava genes. We demonstrate that the interplay between evolutionary conservation and accelerated evolution traces the footprint of genetic load throughout the evolutionary history of cultivated cassava, revealing the impacts of domestication and clonal propagation on gene evolution.

Materials and methods

Sequencing and assembly

We gathered a total of 52 related species in addition to cassava, 27 of which we sequenced and assembled, to evaluate evolutionary conservation and selection across the cassava genome. In order to maximize the amount of evolutionary time sampled; while maintaining reliable alignments to cassava, we sampled 26 species across the Euphorbiaceae family, to which cassava belongs. These species were collected from: the Germplasm Resources Information Network and contributions from many botanic gardens across the United States including the Denver Botanic Garden, the Missouri Botanic Garden, the Montgomery Botanic Garden, the National Botanic Garden, the National Tropical Botanic Garden, The New York Botanic Garden, and the US Botanic Garden.

We then extracted DNA from leaf tissue and sequenced these individuals using Illumina NovaSeq-6000. Genome sizes were estimated using k-mer spectra created using "jellyfish" (Marçais and Kingsford 2011) with a k-value of 21, in order to estimate sequence input coverage for assembly (https://bioinformatics.uconn.edu/genome-size-estimation-tutorial/). Additional short-read sequences were downloaded from the National Center for Biotechnology Information Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/sra/) corresponding to 11 unspecified Euphorbiaceae taxa that were previously part of an effort to digitize a botanic garden (Liu *et al.* 2019). We then used a short-read sequence assembler MEGAHIT (Li *et al.* 2015), with modified parameters of "-m 0.2 --no-mercy --min-count 3 --k-min 31 --k-step 20" to create contig assemblies. These parameters follow recommendations for genome assemblies of complex genomes with >30X sequence coverage (https://github.com/voutcn/megahit).

We additionally obtained long-read sequences using PacBio Sequel II for 7 species among our sampled Euphorbiaceae taxa representing a diverse sample across the family. These include: *Cnidoscolus aconitifolius*, *E. pulcherrima* (poinsettia), *Excoecaria cochinchinensis*, *Garcia nutans*, *Mallotus* sp., *Mercurialis annua*, and *Reutealis trisperma*. These sequences were assembled using Hifiasm (Cheng *et al.* 2021) utilizing default settings. An additional 14 genome assemblies from other related species were downloaded from SRA (https://www.ncbi.nlm.nih.gov/sra/) and added to our assembled genomes resulting in a total of 52 species, excluding cassava (Supplementary Table 1).

Genome quality metrics were calculated to inform their usefulness in later analyses. We calculated and reported assembly size and the length of the shortest contig for which longer and equal length contigs cover at least 50% of the assembly (N50). Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis was performed using the eudicot ortholog lineage database (Simão *et al.* 2015). This metric gives a rough estimate of how well the gene-space is captured by the assembled genome, while also giving a snapshot of the level of gene duplication.

Pan-genome annotation

For the long read assemblies, we performed genome annotation using the BRAKER2 protein homology pipeline (Brůna *et al.* 2021). This BRAKER2 pipeline utilizes ProtHint (Brůna *et al.* 2020) and a protein database consisting of the Viridiplantae clade to produce de novo gene annotations.

We combined these assembled genomes with other Euphorbiaceae public assemblies of *M. esculenta*, *H. brasiliensis*, and *Ricinus communis* to create a pan-genome panel. Orthogroup and synteny analyses were performed using GENESPACE (Lovell *et al.* 2022) using the default pipeline. These orthogroups were used to define homologous genes across cassava for all analyses.

Multiple sequence alignment

To compensate for the large variation in assembly quality, we used a limited alignment process to align fully reconstructed genes in each species. Our methodology followed a multiple sequence pipeline from https://bitbucket.org/bucklerlab/p_reelgene/src/master/ (Schulz *et al.* 2023). Cassava transcripts were aligned to each genome while tracking UTR, intron, and exon positions. Exonic regions in the target genomes that could be aligned by ≥90% of the transcript and had the highest alignment score consolidated with the query transcript. Multiple sequence alignments were created using MAFFT "--ep 0 --genafpair --maxiterate 1000" for each cassava transcript. While this methodology ignores duplicated copies of genes in target genomes, it simplifies analyses by avoiding errors introduced from fragmented assemblies and polyploidy. Additionally, to enable in-frame protein coding analysis across homologous genes, any positions with gaps in the cassava transcript were removed from the multiple sequence alignment.

Gene tree analysis

We performed phylogenetic analyses to assess gene evolution and selection signatures. First, we generated maximum likelihood trees using RAxML "-m GTRGAMMA -p 12345" (Stamatakis 2014) for every transcript with a minimum of 4 aligned genomes. To estimate the neutral evolutionary tree of these species, we randomly sampled 1,000 genes and concatenated the 4-fold degenerate sites (sites where mutations produce no amino acid changes) from their multiple sequence alignments. We rooted this tree to the Malpighiales species *Hypericum perforatum*. This species tree was used only for the visualization of relationships.

Next, we used the phylogenetic analysis by maximum likelihood (PAML) suite of tools to evaluate gene and site level conservation (Yang 2007). For protein coding analysis, we executed 2 different models of the PAML codeml tool. The models differ by either treating the tree as having one ω ratio parameter for the entire tree or allowing 2 ω parameters: one for the cassava branch and one for the rest of the tree. For each test, we used a gene tree generated from the gene of interest. Each model reports the ratio of nonsynonymous to synonymous variants (dN/dS) at each gene as well as the likelihood of the given model. To aid in the interpretability of $dN/dS \gg 1$, we set the maximum $dN/dS = 2$, thresholding all values greater than this to 2. We interpreted the difference between these models as a method of detecting a difference in selection between cassava and the rest of the tree and performed likelihood ratio tests to verify the significance of any differences between these models. Additionally, we performed PAML baseml analysis for each transcript multiple sequence alignment giving an evolutionary rate for every base-pair position. For the subset of pollen-related genes found to be under significant relaxed or positive selection, we analyzed protein evolutions using a PAML Branch-Site model. This model is used to identify which amino acid substitutions are responsible for the differences between our target branch, cassava, and all the other species.

Gene model selection

We used our evolutionary information to filter down the ~64k and ~33k transcript and gene from the CassavaV7.1 annotations, respectively, to 26k gene models with a single most conserved transcript. First, we filtered out any transcripts that had no annotated untranslated regions, as this suggests it has minimal RNA evidence. Next, we retained gene models that were found in at least one of the other 52 assembled species, as de novo gene generation without any homology at this timescale is likely rare. The best transcript for the remaining genes was decided by looking at the lower dN/dS ratio indicating the most evolutionary conservation and potential for functionality. Any remaining multiple transcripts at a gene were filtered to the longest transcript, which, while not necessarily ideal, provides us with a single transcript for future analysis for each of the ~26k genes.

Intraspecific selection signatures

We used a large panel of cassava clones to assess intraspecific measures of selection. From the haplotype map found on cassavabase.org (Fernandez-Pozo *et al.* 2015), we filtered down to 330 cassava clones with at least 10X average genome coverage. We then filtered variant sites to biallelic single nucleotide polymorphisms (SNPs) with <20% missingness and minor alleles having at least 3 occurrences. We calculated the Residual Variation Intolerance Score (RVIS; Petrovski *et al.* 2013), by regressing the number of nonsynonymous SNPs with a minor allele frequency ≥1% onto the total number of SNPs in each gene, then scoring each gene's studentized residual from this regression. To further focus on those functional variations likely to represent deleterious load, we performed the same analysis on nonsynonymous sites flagged as putatively deleterious. These deleterious mutations were classified previously (Long *et al.* 2023) and are sites with a baseml evolutionary rate <0.5, a Sorting Intolerant from Tolerant (SIFT) (Ng and Henikoff 2003) score ≤0.05, and a minor allele frequency <20%. The RVIS methodology was then repeated to form a score we term here as deleterious RVIS (DRVIS).

We also combined interspecific and intraspecific measures of functional variations to test for positive selection. The test performed was the "McDonald–Kreitman test" (McDonald and Kreitman 1991) which can be calculated as $\alpha = 1 - (pNpS/dN/dS)$, where pNpS is the ratio of nonsynonymous mutation sites to synonymous mutation sites within a population. We used the cassava hapMap (Ramu *et al.* 2017) population to determine pNpS, using a minor allele frequency cutoff of 10%, and calculated α. Similar enrichment tests to those previously described were performed with α, where the $\alpha > 0$ describes the proportion of sites fixed by positive selection (Supplementary Table 3 and Fig. 3).

Selection evaluation

We collected gene ontologies (GOs) through homology to the TAIR10 *Arabidopsis thaliana* genes (https://www.arabidopsis.org/). BLASTP was performed between CassavaV7 and TAIR10 to determine homologous genes. GO term enrichment was performed using the "topGO" package in R (Version 2.50.0), analyzing GO terms for biological processes in regard to each of our selection measures (Tables 1–3, Supplementary Table 3). GO terms were filtered to only include terms that are present in more than 1 gene. GO term significance is the result of a Fisher exact test with the associated *P*-value. The top 10 GO terms for each test were reported, and with a total of 5,828 GO terms tests, a Bonferroni multiple test correction *P*-value threshold of 8.57e−6 can be used to consider significance. We used public databases of *A. thaliana* gene

Table 1. Enriched GO terms for genes under relaxed selection.

GO.ID	Term	Annotated	Significant	Expected	P-value
GO:0009860	**Pollen tube growth**	177	6	0.34	5.1e−06
GO:0060321	**Acceptance of pollen**	12	2	0.02	0.00023
GO:0006893	Golgi to plasma membrane transport	27	2	0.05	0.00122
GO:0009409	Response to cold	652	6	1.25	0.00143
GO:0072583	Clathrin-dependent endocytosis	46	2	0.09	0.00350
GO:0009651	Response to salt stress	940	7	1.80	0.00495
GO:0006900	Vesicle budding from membrane	64	2	0.12	0.00667
GO:0009846	**Pollen germination**	79	2	0.15	0.01001
GO:0046777	Protein autophosphorylation	264	3	0.50	0.01397
GO:0010224	Response to UV-B	108	2	0.21	0.01814

GO term enrichment produced from "topGO" among genes significantly relaxed from selection ($\Delta dN/dS < 0$ and Bonferroni multiple test correction significance). Pollen related GO terms are shown in bold.

Table 2. Top 10 enriched GO terms for genes with excessive nonsynonymous mutations (RVIS genes).

GO.ID	Term	Annotated	Significant	Expected	P-value
GO:0006952	Defense response	3,176	250	138.86	6.0e−15
GO:0007165	Signal transduction	3,162	186	138.24	1.2e−10
GO:0006468	Protein phosphorylation	1,473	118	64.40	1.7e−09
GO:1902065	Response to L-glutamate	19	9	0.83	3.5e−08
GO:0042742	Defense response to bacterium	1,308	101	57.19	2.7e−07
GO:0007166	Cell surface receptor signaling pathway	71	15	3.10	3.4e−07
GO:0050832	Defense response to fungus	831	64	36.33	6.9e−06
GO:0010483	**Pollen tube reception**	26	8	1.14	1.0e−05
GO:0032922	Circadian regulation of gene expression	35	9	1.53	1.4e−05
GO:0009616	RNAi-mediated antiviral immune response	28	8	1.22	1.8e−05

GO term enrichment produced from "topGO" among genes in the top 5% of RVIS scores. Pollen related GO terms are shown in bold.

expression atlases (Waese *et al.* 2017, https://bar.utoronto.ca/eplant/) to compare tissue-specific expression of orthologous genes to the set of 48 genes with relaxed selection signatures (Supplementary Table 2; Nakabayashi *et al.* 2005; Schmid *et al.* 2005).

Cassava underwent a paleopolyploidy event and many genes contain duplicates throughout its genome. Because of this genome duplication, we addressed our measures of selection on individual genes as well as consolidating their metrics by ortholog group, by recording the least extreme value (i.e. smallest absolute value or least significant *P*-value). Additionally, we analyzed differences in selection signatures between paralogues within ortholog groups (Supplementary Fig. 6) and measured evolutionary distances to each assembled species and known homeologous chromosome pairs (Bredeson *et al.* 2016) to determine if there is any evidence for allopolyploidization, which would show asymmetric similarity to a relative of a putative diploid progenitor (Supplementary Fig. 4). Additionally, we binned our selection measurements (dN/dS, RVIS, DRVIS, etc.) and other genome annotations (gene density, genetic map, and domestication sweeps; Ramu *et al.* 2017) into 250 kb bins to examine large-scale signatures of asymmetric selection across homeologous chromosomes (Supplementary Fig. 5).

Differential expression

Differential expression between flower (mixed female and male inflorescences) and nonflower tissues was performed for the 48 genes found to be under relaxed selection. RNA sequence counts across 5 different tissues and 150 cassava clones from a previous study were used to evaluate gene expression (Ogbonna *et al.* 2021). Differential expression analysis was performed using R package "DEseq2" (Love *et al.* 2014), comparing flower tissue expression to all other tissues. This analysis compares the 2 sets of expression data, flower tissue and nonflower tissue, and reports the Log2Fold changes in expression and the associated *P*-value for each gene. This was used to help support the floral function of the annotated genes. Log2Fold changes in expression with significance levels were reported (Supplementary Fig. 1).

Results

Genome assemblies

In conjunction with a breeding study, we sequenced and assembled 27 genomes for the purpose of training a machine-learning model to improve genome-wide selection models for breeding (Long *et al.* 2023). In this study, we annotate and characterize these genomes to compare ortholog conservation across the cassava genome. Most of these species were sequenced using short-reads, while 7 species were sequenced using long-read sequencing. In addition to these sampled species, we assembled 11 Euphorbiaceae taxa with publicly available short-read sequences from a botanic garden survey (Liu *et al.* 2019), and collected 15 public reference assemblies (Tuskan *et al.* 2006; Bredeson *et al.* 2016; Xu *et al.* 2017; Horvath *et al.* 2018; Chen, Huang, *et al.* 2019; Chen, VanBuren, *et al.* 2019; Jalali *et al.* 2020; Liu *et al.* 2020; Wei *et al.* 2020; Wu *et al.* 2020; Zhou *et al.* 2020; Cai *et al.* 2021; He *et al.* 2021; Zhou *et al.* 2021; Lu *et al.* 2022) for a total of 53 species (Supplementary Table 1). Genome sizes and sequence coverage were estimated through k-mer analysis (Supplementary Table 1).

Genome assembly quality was evaluated by assembly size, contiguity, and reconstructed gene-space (Fig. 1). The quality of gene-space reconstruction was estimated through BUSCO, and contiguity quality represented by the length of the contig of which 50% of the assembly is contained with that size of contig or larger (N50). These assembled genomes have large variability in contiguity and quality of gene-space reconstruction due to differences in sequencing methods as well as large variability in genome size (Fig. 1). Species assembled from long-reads are of very high quality

Table 3. Top 10 enriched GO terms for genes with a buildup of deleterious mutations (DRVIS genes).

GO.ID	Term	Annotated	Significant	Expected	P-value
GO:0006468	Protein phosphorylation	1,473	144	75.04	2.6e−19
GO:0080168	Abscisic acid transport	24	13	1.22	2.1e−11
GO:0048544	**Recognition of pollen**	100	25	5.09	2.2e−11
GO:0090332	Stomatal closure	99	17	5.04	1.9e−06
GO:0010496	Intercellular transport	43	9	2.19	3.2e−06
GO:0015692	Lead ion transport	19	7	0.97	2.5e−05
GO:0019318	Hexose metabolic process	114	13	5.81	3.6e−05
GO:0007165	Signal transduction	3,162	197	161.09	4.8e−05
GO:0090436	Leaf pavement cell development	10	5	0.51	6.9e−05
GO:0051645	Golgi localization	10	5	0.51	6.9e−05

GO term enrichment produced from "topGO" among genes in the top 5% of DRVIS scores. Pollen related GO terms are shown in bold.

with N50 and BUSCO values comparable or higher to many of the previously published reference assemblies.

Pan-genome annotations

We combined our long-read assemblies with publicly available Euphorbiaceae genome assemblies to create a Euphorbiaceae gene pan-genome. Our de novo assemblies were annotated using BRAKER2, and genome homology and synteny were produced through the tool GENESPACE (Lovell *et al.* 2022). This pan-genome defines orthogroups for each gene, including cassava genes. There are over 11k orthogroups that are found in at least 80% of high-quality Euphorbiaceae assemblies in our pan-genome (Supplementary Fig. 1). These conserved orthogroups account for ∼19k cassava genes (72% of the high-quality genes) and are likely biased toward syntenic genes.

Evolutionary conservation across the Euphorbiaceae family

We measured the presence of reconstructed gene orthologs across our phylogeny and found a wide distribution of how many genome assemblies had complete homologous sequence for each cassava gene (Fig. 2a). We considered alignments for the single best-aligned homologous gene from each genome assembly with at least a 90% aligned length to the cassava homolog. While the distribution of taxa with homologous genes is dependent on assembly quality, there are many genes that are present and completely assembled across a majority of Euphorbiaceae and related species, even from the short-read assemblies. We see very few genes across all 53 species, likely hindered by poor assembly and alignment quality derived from short-read assemblies. The set of cassava genes with very few observations across the Euphorbiaceae may indicate those genes are either unique to cassava, conserved in few species, or may be nonfunctional annotations thus not conserved.

We constructed a phylogeny to evaluate relationships among these 53 species. A maximum likelihood neutral tree was estimated from 4-fold degenerate sites in a random sample of 1,000 genes (Fig. 2b). This phylogeny shows relationships that agree with previously understood taxonomic relationships, including 3 subfamilies: Acalyphoideae, Crotonoideae, and Euphorbioideae (Wurdack *et al.* 2005). Two species, *Breynia nivosa* and *Flueggea neowawraea*, were previously classified as biovulate subfamily Phyllanthoideae, and are now part of a separate family Phyllanthaceae (Wurdack *et al.* 2005). Outgroup species, including aspen, willow, and flax, are distantly related but still fall within the order of Malpighiales.

While the contiguity and quality of short-read assemblies are relatively low (Fig. 1a), their assembly of genic regions allowed us to incorporate these species into evolutionary assessment. Many of the short-read assemblies from genomes that were ≤1 Gbp in size had high contiguity and BUSCO scores, while those from species with larger genome sizes are of lower quality. This is mainly due to common sources of difficulty such as obtaining high enough coverage sequence data and assembling large complex regions with short-read information. Ultimately the utility of these genomes is visible in the amount of reconstructed ortholog space in the cassava genome (Fig. 2a).

Interspecific selection signatures

Using evolutionary conservation and interspecific variation, we analyzed selection signatures across ∼26k cassava gene models that passed quality filters. We calculated the ratio of nonsynonymous, causing amino acid changes in protein sequence, to synonymous, putatively neutral, substitutions across the evolutionary tree (dN/dS, Kryazhimskiy *et al.* 2008, Fig. 3a). Additionally, we estimated this ratio for substitutions occurring along the cassava branch of the tree. Comparing nucleotide sequence conservation across millions of years allowed us to measure the selection on cassava genes. Our results show the majority of genes in the cassava genome have a dN/dS < 0.5 implying purifying selection (Fig. 3a). These genes are likely functional and important to be conserved across the Euphorbiaceae family and related species, and to have low tolerance for mutations that disrupt conserved function. Genes with dN/dS ∼ 1 are likely not under strong selection and are either nonfunctional or whose function is not currently beneficial. Genes with dN/dS > 1 are either under positive selection, relaxed purifying selection, pseudogenes, or are poorly estimated due to short gene length (Fig. 3a). For visual clarity we truncated all dN/dS ratios >2, setting their value to 2. Using the PAML branch test, we tested the relaxation of selection in cassava and reported the difference between the dN/dS of the evolutionary tree and that of the cassava branch. The negative ΔdN/dS indicates a larger dN/dS value in the cassava branch of the tree, or a transition away from purifying selection. By comparing this branch-specific ratio to the value estimated across all species we found 167 genes comprising 148 different orthogroups that showed significantly higher dN/dS values along the cassava lineage (Supplementary Table 2), suggesting relaxed or positive selection in cassava (Fig. 3b). We found that the reliability of these estimates of dN/dS and the significance of separations between cassava and the other species is dependent on the length of the coding sequence and the number of species with reconstructed orthologs, with several short genes experiencing large, but insignificant differences in dN/dS ratios (Fig. 3b and c—lower left quadrants). This results from limited observations of the

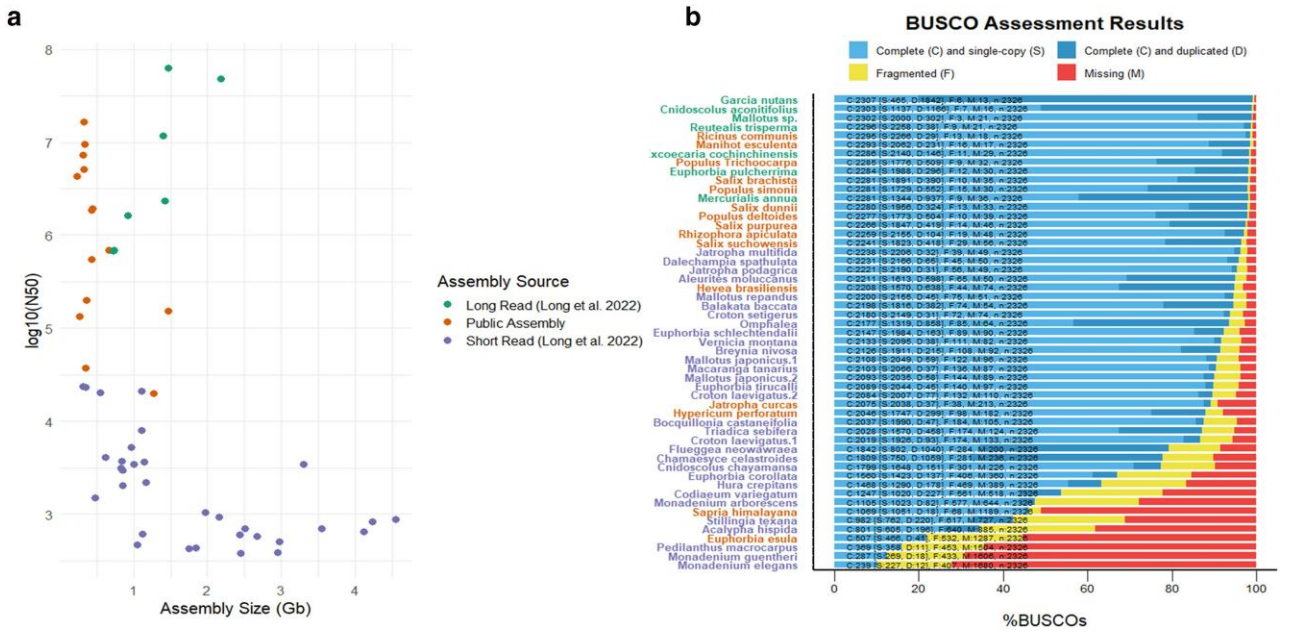

Fig. 1. Assembly quality statistics. Assembly size and N50 are shown for long and short-read assemblies we produced, as well as the public assemblies used in this study (a). BUSCO scores are plotted with species' text color matching sources (b).

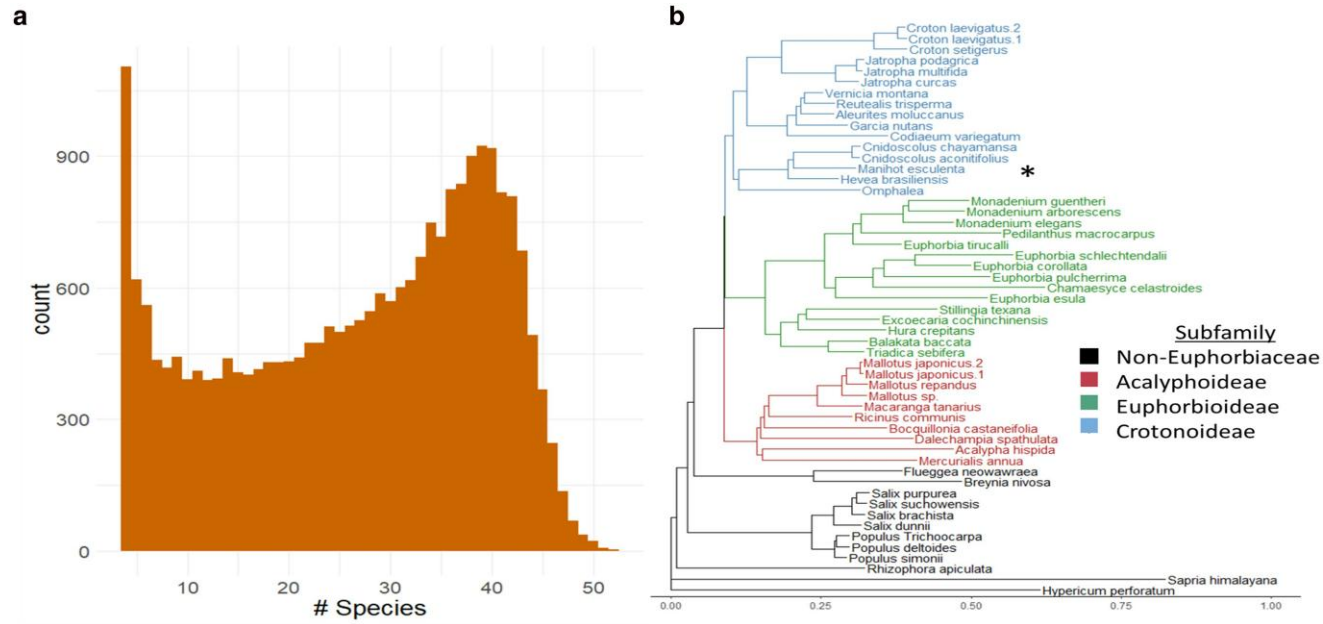

Fig. 2. Ortholog occurrence across all assemblies and phylogenetic relationships. An ortholog frequency histogram with the number of species that are represented in each ortholog group across all assemblies (a). Phylogenetic tree created from 4-fold degenerate sites from 1,000 randomly selected genes. The Euphorbiaceae subfamilies are designated by color (b). Cassava (*M. esculenta*) is part of the Crotonoideae subfamily and designated with "*".

ratio of nonsynonymous to synonymous mutations that can result in extremes that are not statistically relevant.

The ancestor of cassava experienced a whole genome duplication ~40 million years ago (Bredeson *et al.* 2016), and genes resistant to fractionation have been retained as duplicated genes across the genome. Duplicate genes may show signatures of relaxed selection, if one copy maintains function and the other subfunctionalizes or neofunctionalizes (Flagel and Wendel 2009). To minimize the possibility of conflating ongoing fractionation with relaxed selection on ancestral function, we only considered orthologous groups of genes that contained either a single cassava gene or where all cassava gene copies passed significance tests, resulting in 48 genes from 47 orthogroups (Fig. 3c).

Since its domestication, selection in cassava has been strongest on root mass. We expected traits unnecessary for the clonal reproduction of these large roots such as photosensitivity, flowering and seed production, and perennialism to be released from selection. We performed differential gene expression using available RNA-sequencing data from 5 cassava tissues and found that from among the 48 genes showing relaxed selection, 15 had differentially increased expression in flowers compared to nonflower tissues (Supplementary Fig. 2). Additionally, ~70% of the

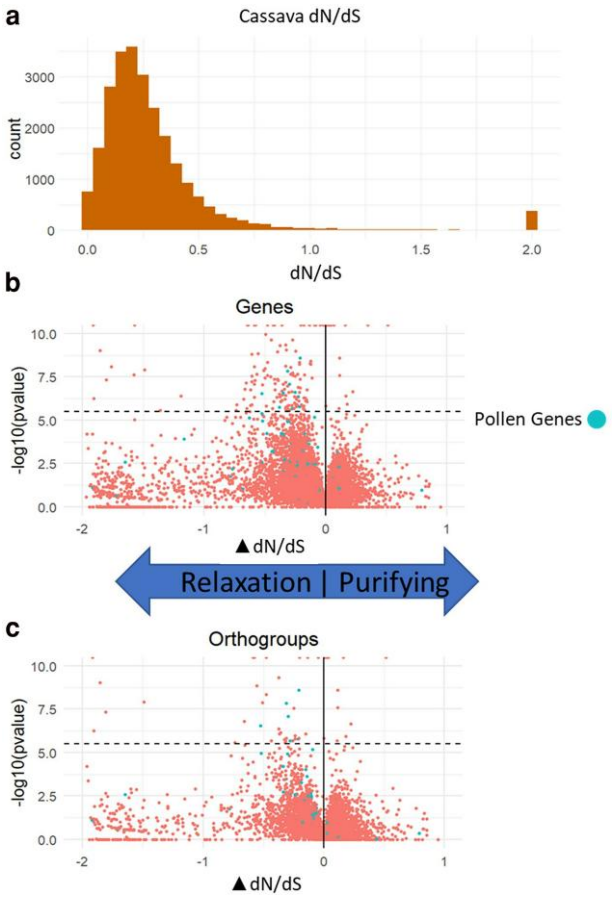

Fig. 3. Selection signatures dN/dS gene conservation. a) Histogram of dN/dS values from all genes across cassava, with values >2 plotted at dN/dS = 2 (top). b) The difference in dN/dS score between the 52 species used in this study and cassava for each gene in cassava, with the y-axis showing the log ratio test *P*-value between these 2 models and dotted line showing multiple test correction significance threshold. c) The difference in dN/dS score between the 52 species used in this study and cassava summarized for each orthogroup in cassava, with the y-axis showing the log ratio test *P*-value between these 2 models and dotted line showing multiple test correction significance threshold. Arrows indicate difference in selection in cassava (i.e. ΔdN/dS < 0 implies a relaxation of purifying selection in cassava, or a transition to more positive selection, and ΔdN/dS > 0 implies a stronger purifying selection in cassava).

A. thaliana homologs to the 48 genes are most highly expressed in flower, seed, and pollen tissues, (Supplementary Table 2).

We then investigated the possible biological functions of this set of genes that show relaxation from purifying selection in cassava compared to the rest of the evolutionary tree. We performed GO enrichment for GO terms regarding biological processes. The set of 48 genes with significant differences in dN/dS values showed an enrichment for processes involved with pollen and pollen tube development, which exhibited a 20-fold enrichment (Table 1). These ontologies were attributed to 6 specific genes in cassava: Manes.02G178800, Manes.03G204900, Manes.03G130950, Manes.04G017000, Manes.04G056400, and Manes.08G062900, all were expressed in floral tissues and 4 were expressed differentially higher in flowers relative to leaf, stem, fibrous, and storage root tissues. While less pronounced other showed significant enrichment in the set of relaxed genes including responses to multiple stresses (cold, salt, UV-B). These may be the result of domestication and the removal of stresses that exist in the species' wild, perennial state.

Intraspecific selection signatures

Relaxed selection along the cassava lineage identified candidate genes, but it is unclear whether the relaxation of selection occurred across the genus *Manihot*, or after domestication. To determine which pathways were released from selection in domesticated cassava, we used intraspecific data with a sample of 330 sequenced cassava clones. We used 2 versions of the residual variation intolerance score (RVIS, Petrovski *et al.* 2013), which uses either excess nonsynonymous (Fig. 4a) or deleterious mutations (Fig. 4b) to identify outlier genes. RVIS statistics help control for protein length and differential mutation and drift at the gene level. To identify genes with extremes of deleterious variation, we calculated a DRVIS score using variant sites classified as deleterious through evolutionary conservation (DRVIS, Fig. 4b). The genes in the top 5% ($n = 1287$) of RVIS (Table 2) and DRVIS (Table 3) scores, representing those genes with high polymorphism at functional sites, also showed enrichment for pollen-related biological processes, among other biological processes including multiple processes responsible for defense against biotic stresses.

Given that pollen-related traits are enriched among the extremes of both interspecific and intraspecific measures of genetic load, we further investigated all 348 genes that had pollen-related GO terms, irrespective of whether they reached significance in individual tests. We performed χ^2 tests for significant differences in distributions of ΔdN/dS, RVIS, and DRVIS between 348 pollen-related genes and all other cassava genes (Fig. 5). We found ΔdN/dS (*P*-value = 0.027) and RVIS (*P*-value = 0.0055) to be significantly lower in pollen-related genes than all other genes, while DRVIS (*P*-value = 3.3e−05) was significantly higher. All these effects, while significant, are small, but are consistent with relaxation from selection among pollen genes in cassava.

In addition to dN/dS, RVIS, and DRVIS we also used the MK test to evaluate selection measures. The MK test uses a combination of intraspecific and interspecific functional variation to measure selection with a positive value ($\alpha > 0$) indicating the proportion of substitutions fixed by positive selection. GO term enrichment was performed on genes in the top 5% of the MK test α value ($\alpha = 1$), and found many functions related to plant defense, like RVIS (Supplementary Table 3). The MK test α value also showed enrichment for pollen tube functions, though overall no significant difference across all pollen genes (Supplementary Fig. 2). Negative estimates of α are common in plant populations (Gossmann *et al.* 2010), as we observe for the majority of genes in cassava. This may be due to low effective population sizes in cassava, intensified by the population contractions seen during the domestication and improvement bottlenecks (Ramu *et al.* 2017), altering the landscape of segregating variation. Additionally, MK tests are also sensitive to segregating slightly deleterious variation (Messer and Petrov 2013), limiting the ability to detect positive selection.

Chromosome evolution

Knowing that the cassava genome has experienced a paleotetraploidy event, we examined previously characterized homeologous chromosomes to look for any asymmetrical measures of conservation and selection. We took the average genetic distance to each other genome assembly across all cassava genes. These distances were then compared across homeologous chromosomes to look for any evidence of chromosome lineage resulting from an allopolyploid event (Supplementary Fig. 4). We found that with the current resolution provided from the genome assemblies in this

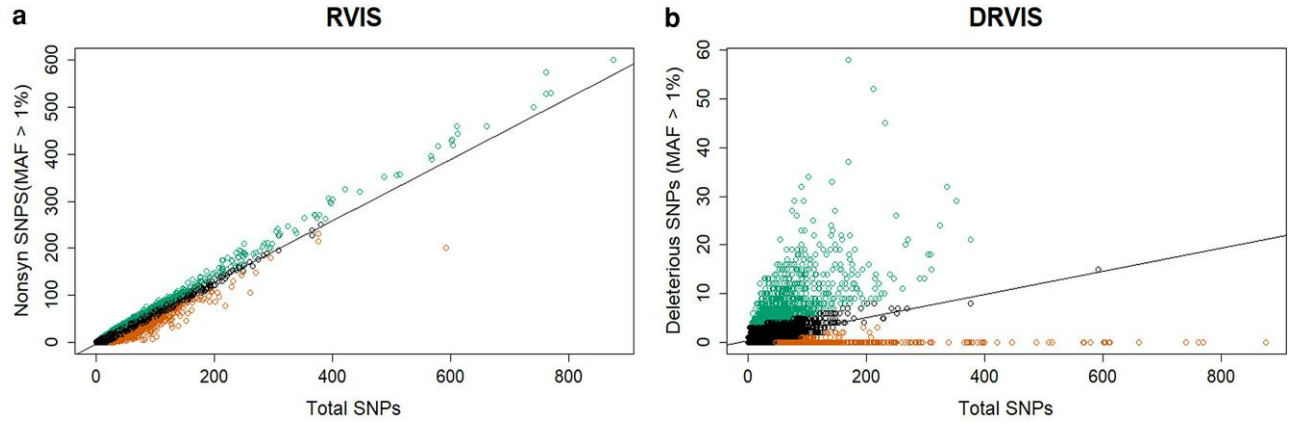

Fig. 4. Identification of deleterious mutations in genes using within species residual variation intolerance scores. Regression for number of nonsynonymous SNPs (a) and putative deleterious SNPs (b) against the total number of SNPs in each gene in cassava. The residuals from each regression give RVIS and DRVIS scores, with the top and bottom 5% ($n = 1287$) shown.

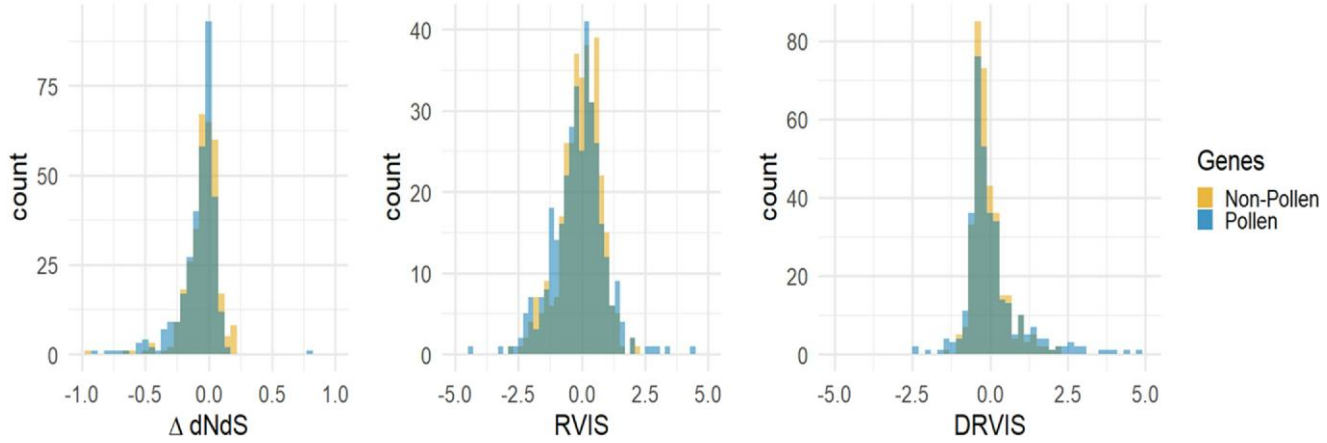

Fig. 5. Distributions of ΔdN/dS, RVIS, and DRVIS between pollen and nonpollen-related genes. Histograms are shown between pollen (blue) and nonpollen (orange) related genes for ΔdN/dS, RVIS, and DRVIS. Nonpollen-related genes are subsampled to an equal number of genes for visual comparison.

study, there is no evidence for allopolyploidization; however, this cannot be certain from the available information.

We also examined our selection metrics across the cassava chromosomes to look for any signatures of biased fractionation, due to one homeologous chromosome being selectively conserved over the other (Supplementary Fig. 5). While regions of some chromosomes show some signatures of decay (low gene densities, low recombination rates, high dN/dS values), there does not appear to be any evidence for whole chromosome degradation. However, among genes with multiple paralogues, there is evidence for gene subfunctionalization or degradation as seen by one gene copy showing high conservation over other gene copies (Supplementary Fig. 6).

Discussion

The intricate interplay between natural selection, domestication, and prolonged clonal propagation has shaped the cassava genome over the course of its cultivated history. We explore the impact of clonal propagation after domestication, and the absence of recombination, has had on the accumulation of deleterious load in the cassava genome. Employing interspecific and intraspecific analyses, we evaluate selection signatures across the genome, corroborate with tissue-specific expression, and investigate affected gene pathways, and finding evidence for relaxed selection in pollen and flower-related genes.

Selection signals from comparative genomics

Surveying millions of years of evolution by sampling taxa across the Euphorbiaceae, we measured selection on each cassava gene. Most genes under differential selection are relaxed along the cassava lineage (Fig. 3b and c) and are enriched for pollen-related functions. Sexual-related genes have been observed to experience faster evolutionary rates in general (Cutter and Ward 2005), and these genes may be predisposed for faster mutation accumulation. Asexual reproduction may allow for the accumulation of increased mutational load in sexual function genes, as these genes are likely no longer required for effective reproduction. Related breakdowns in genes involved in pollen function have been observed between outcrossing wild species and clonally propagated cultivars, for example in potato (Hardigan *et al.* 2017) and the ornamental *Ranunculus* genus (Kocot *et al.* 2022).

Even wild species may show signatures of this process. In the tree species *Populus tremuloides*, male fertility declines with the clonal age of an individual, potentially due to the accumulation of somatic mutations (Ally *et al.* 2010). In *Decadon verticillatus*, a transition to asexual reproduction led to the loss of sexual compatibility, primarily through pollen dysfunction (Eckert *et al.* 1999).

Comparative evolutionary signal supports a role in the sexual reproduction of the 48 relaxed cassava genes. Two of these genes, Manes.02G178800 and Manes.08G062900 are annotated as exocyst and secretory complexes that have been experimentally shown to be necessary for proper pollen development in *A. thaliana* (Marković *et al.* 2020). Two more genes, Manes.03G130950 and Manes.04G017000, are annotated as members of the ATP-binding cassette transporter family, whose homologs are essential for anther and pollen exine development in rice (Qin *et al.* 2013).

Intraspecific selection signals

Genes under relaxed selection in the reference genome of *M. esculenta* relative to other taxa could represent episodes of selection that happened during speciation, or during domestication. To disentangle these effects, we used genotyped cassava clones to provide a within species perspective on selection. Similar to the evolutionary signal seen in comparative genomics across the Euphorbiaceae, genes with pollen-related functions showed enrichment for functional variation, measured by RVIS and DRVIS, between cassava clones. Further analysis of RNA expression data in cassava supports the sexual reproduction-related functions of the 48 genes that showed relaxed selection, with many of them showing differentially increased expression in flower tissues (Supplementary Fig. 2).

While GO term enrichment may be prone to overinterpretation, it is a common method used to identify possible affected biological processes and combined with other analyses can be one signal supporting a conclusion. Other significantly enriched GO terms from these population-level analyses included defense response functions. This may result from disruptive positive selection, as high amounts of functional variation can be beneficial. Selection for the high diversity of plant defense genes has been previously shown in plants (Zhang *et al.* 2014, 2016), and these defense response genes may be relevant for common diseases afflicting cultivated cassava such as cassava mosaic virus or cassava brown streak disease. Genes with functions related to circadian rhythms also showed evidence for positive selection, which agrees with previous understanding of gene functions in plants commonly under positive selection (Michael *et al.* 2003). None of our analyses resulted in pathways involved in root development or function, even though large amounts of selection have been applied to this trait. This may be because GO terms derived primarily from studies of Arabidopsis may not have a direct correlation to pathways responsible for the development of large storage roots like those found in cassava. While GO terms are a coarse estimation for function, these and the other significant functional elements may be further investigated to shed light on molecular functions behind cassava evolution, domestication, and cultivation.

Sexual recombination and seed production are essential to combine favorable alleles to create improved cultivated varieties. Low fruit and seed production hinders breeding efforts in cassava. Low rates of female flower production and large variation in flowering times have been targets to alter for increased cassava seed production (Oluwasanya *et al.* 2021). Flowering induction is only part of the problem, however, as many studies have reported flower abortion rates of over 80% (Ukwu *et al.* 2018; Ramosy Abril *et al.* 2019; Bandeira E Sousa *et al.* 2021). Studies have also found genotypic variability in pollen viability in cassava, and that self-incompatibility does not explain this variability (Ramos Abril *et al.* 2019; Bandeira E Sousa *et al.* 2021). It has been suggested that fruit abortion in outcrossing species may be due to deleterious mutations (Wiens *et al.* 1987). Multiple studies have shown low pollen amounts and low pollination rates in cultivated cassava crosses compared to wild progenitors and other Manihot species (da Jennings 1963; Vieira *et al.* 2012; Silva *et al.* 2018). Previous studies on the relationship between cassava clonality on deleterious mutations have shown an unexpected lack of correlation between recombination and deleterious mutations (Ramu *et al.* 2017), supporting the conclusion that these mutations are indeed being enriched through absent recombination from clonal reproduction.

Degradation of sexual function genes

The results from both interspecific evolutionary analyses as well as intraspecific variation support the conclusion that genes responsible for sexual reproduction, specifically pollen function, have experienced disproportionate mutation accumulation. Domestication can increase the frequency of deleterious mutations, through both hitchhiking and the increased genetic drift of a bottlenecked population. Extended periods of clonal propagation can also reduce the ability to effectively purge deleterious mutations due to the lack of independent assortment and recombination (Muller 1964; Hill and Robertson 1966). Disentangling the relative contributions of domestication and clonal propagation is difficult in clonally propagated crops, as the 2 processes intertwine, and we will be unlikely to be able to fully attribute deleterious mutation accumulation to either process. Unlike other species like maize (Rodgers-Melnick *et al.* 2015) and humans (Hussin *et al.* 2015) where deleterious mutations are enriched in low recombination regions of the genome, in cassava, deleterious mutation abundance and recombination frequency show little correlation (Ramu *et al.* 2017). This suggests that the deleterious alleles that have reached fixation in cassava have done so on large haplotypic backgrounds, as expected with low recombination occurrence due to clonal reproduction. However, as recombination is not absent, cassava populations do not fit classical models of Hill–Robertson interference and Muller's ratchet (Felsenstein 1974; Charlesworth and Charlesworth 1997). Some cassava individuals maintain 2 divergent haplotypes, as may be expected with asexuality, but many do not (Qi *et al.* 2022).

Alongside this general accumulation of deleterious mutations, the disproportionate accumulation in genes related to sexual reproduction may be caused by relaxed selection on their function. Several domesticated clonal crop species have lost their sexual fitness (Mckey *et al.* 2010). One potential explanation of deceased sexual fitness is counter-selection against sex, where lower investment in sexual reproduction may allow for more energy toward yield-related traits as in potato (Simmonds 1997; Mckey *et al.* 2010). In cassava, selection against floral development may prevent branching, as branching precedes the development of flowers, but is an undesirable trait in agronomic contexts (Pineda *et al.* 2020). The decrease in population size from a domestication bottleneck can cause deleterious mutations to rise to high frequency (Bosse *et al.* 2019), and while domestication has also been linked to selection in flowering genes, especially in regard to photoperiod sensitivity (Lin *et al.* 2021), this may be less

impactful for crop species such as cassava whose agronomic value is not derived from fruit or seed. The putative accelerated accumulation of mutations in sexually related genes may offer a snapshot of where in the genome these fitness impacts are experienced at higher rates.

Conclusion

This work has produced a deep evolutionary resource for the evaluation of selection and deleterious mutations in cassava. Evolutionary conservation across the Euphorbiaceae family can help determine the functional importance of genes across the cassava genome. Since its domestication and transition to clonal propagation, cassava had an accumulation of deleterious mutations and a loss of overall fitness, especially regarding sexual reproductive fitness. Understanding the impacts of domestication and clonal propagation on genetic load in genes related to sexual reproduction can help overcome the reproductive hurdles in cassava breeding and may provide targets for gene editing to repair haplotypes otherwise beneficial to tuber traits necessary for cassava's economic value Fernandez-Pozo *et al.* 2015. These results address only one aspect of genetic load and deleterious mutations in cassava, but the evolutionary resource produced has the potential to address many more in the future. Understanding which genes are under selection pressure in cultivated cassava may help prioritize targeted genetic interventions for those genes that are likely functional and impactful. Genes responsible for sexual reproduction may serve as a target of breeding or genetic manipulation to maintain sexual fecundity for the continual breeding ability of cassava lines.

Data availability

The inputs to analyses, code to reproduce tables and plots, and summary tables can all be found on the GitHub repository (https://github.com/em255/CassavaEuphorbiaceaeGeneEvolution). All sequence and assembly data used in this study were previously published (Long *et al.* 2023). Euphorbiaceae sequence reads and assemblies generated in this study are available under bioprojects PRJNA608937 on the Sequence Read Archives and PRJEB55682 on the European Nucleotide Archive, respectively.

Supplemental material available at G3 online.

Acknowledgments

We would like to acknowledge the many germplasm sources that contributed tissue for sequencing (Supplementary Table 1) including: the Denver Botanic Garden, Germplasm Resources Information Network, the Missouri Botanic Garden, the Montgomery Botanic Garden, the National Botanic Garden, the National Tropical Botanic Garden, The New York Botanic Garden, and the US Botanic Garden. Their support was essential in sampling the vast number of species used in this study. We would also like to thank the NextGen Cassava community for the contribution of cassava genotype data that we used from cassavabase.org (Fernandez-Pozo *et al.* 2015), as well as the National Center for Biotechnology Information Sequence Read Archive (Katz *et al.* 2022) for maintaining access to genome sequences.

Funding

This work is supported by workforce development fellowship Project: NYC-149949, Award: 2021-67034-34970 from the United States Department of Agriculture National Institute of Food and Agriculture as well as start-up funds from the Robbins lab at Cornell. M.C.S. was supported by Nation Science Foundation Postdoctoral Research Fellowship in Biology No. 1907343. A.J.S. was supported by NSF Graduate Research Fellowship DGE—2139899. The authors thank the UK's Foreign, Commonwealth & Development Office (FCDO) and the Bill & Melinda Gates Foundation (Grant INV-007637 http://www.gatesfoundation.org) for their financial support. Additionally, this study is made possible by the funding and support of the United States Department of Agriculture-Agricultural Research Service.

Conflicts of interest

The author(s) declare no conflicts of interest.

Literature cited

Ally D, Ritland K, Otto SP. 2010. Aging in a long-lived clonal tree. PLoS Biol. 8(8):19–20. doi:10.1371/journal.pbio.1000454.

Bandeira E Sousa M, Andrade LRB, Souza EH, Alves AAC, de Oliveira EJ. 2021. Reproductive barriers in cassava: factors and implications for genetic improvement. PLoS One. 16(11):e0260576. doi:10.1371/journal.pone.0260576.

Bosse M, Megens HJ, Derks MFL, de Cara ÁMR, Groenen MAM. 2019. Deleterious alleles in the context of domestication, inbreeding, and selection. Evol Appl. 12(1):6–17. doi:10.1111/eva.12691.

Bredeson JV, Lyons JB, Prochnik SE, Wu GA, Ha CM, Edsinger-gonzales E, Grimwood J, Schmutz J, Rabbi IY, Egesi C, *et al.* 2016. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. Nat Biotechnol. 34(5):562–570. doi:10.1038/nbt.3535.

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom Bioinform. 3(1):lqaa108. doi:10.1093/nargab/lqaa108.

Brůna T, Lomsadze A, Borodovsky M. 2020. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genom Bioinform. 2(2):lqaa026. doi:10.1093/nargab/lqaa026.

Cai L, Arnold BJ, Xi Z, Khost DE, Patel N, Hartmann CB, Manickam S, Sasirat S, Nikolov LA, Mathews S, *et al.* 2021. Deeply altered genome architecture in the endoparasitic flowering plant *Sapria himalayana* Griff. (Rafflesiaceae). Curr Biol. 31(5):1002–1011.e9. doi:10.1016/j.cub.2020.12.045.

Ceballos H, Iglesias CA, Pérez JC, Dixon AGO. 2004. Cassava breeding: opportunities and challenges. Plant Mol Biol. 56(4):503–516. doi:10.1007/s11103-004-5010-5.

Charlesworth B, Charlesworth D. 1997. Rapid fixation of deleterious alleles can be caused by Muller's ratchet. Genet Res (Camb). 70(1):63–73. doi:10.1017/S0016672397002899.

Chen JH, Huang Y, Brachi B, Yun QZ, Zhang W, Lu W, Li HN, Li WQ, Sun XD, Wang GY, *et al.* 2019. Genome-wide analysis of Cushion willow provides insights into alpine plant divergence in a biodiversity hotspot. Nat Commun. 10(1):5230. doi:10.1038/s41467-019-13128-y.

Chen L-Y, VanBuren R, Paris M, Zhou H, Zhang X, Wai CM, Yan H, Chen S, Alonge M, Ramakrishnan S, *et al.* 2019. The *bracteatus* pineapple genome and domestication of clonally propagated crops. Nat Genet. 51(10):1549–1558. doi:10.1038/s41588-019-0506-8.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 18(2):170–175. doi:10.1038/s41592-020-01056-5.

Cutter AD, Ward S. 2005. Sexual and temporal dynamics of molecular evolution in *C. elegans* development. Mol Biol Evol. 22(1):178–188. doi:10.1093/molbev/msh267.

de Freitas JPX, da Silva Santos V, de Oliveira EJ. 2016. Inbreeding depression in cassava for productive traits. Euphytica. 209(1):137–145. doi:10.1007/s10681-016-1649-7.

Dwivedi SL, Heslop-Harrison P, Spillane C, McKeown PC, Edwards D, Goldman I, Ortiz R. 2023. Evolutionary dynamics and adaptive benefits of deleterious mutations in crop gene pools. Trends Plant Sci. 28(6):685–697. doi:10.1016/j.tplants.2023.01.006.

Eckert CG, Dorken ME, Mitchell SA. 1999. Loss of sex in clonal populations of a flowering plant, *Decodon verticillatus* (Lythraceae). Evolution. 53(4):1079–1092. doi:10.2307/2640813.

Felsenstein J. 1974. The evolutionary advantage of recombination. Genetics. 78(2):737–756. doi:10.1093/genetics/78.2.737.

Ferguson ME, Shah T, Kulakow P, Ceballos H. 2019. A global overview of cassava genetic diversity. PLoS One. 14(11):e0224763. doi:10.1371/journal.pone.0224763.

Fernandez-Pozo N, Menda N, Edwards JD, Saha S, Tecle IY, Strickler SR, Bombarely A, Fisher-York T, Pujar A, Foerster H, *et al.* 2015. The Sol Genomics Network (SGN)–from genotype to phenotype to breeding. Nucleic Acids Res. 43(Database issue):D1036–D1041. doi:10.1093/nar/gku1195.

Flagel LE, Wendel JF. 2009. Gene duplication and evolutionary novelty in plants. New Phytol. 183(3):557–564. doi:10.1111/j.1469-8137.2009.02923.x.

Gossmann TI, Song BH, Windsor AJ, Mitchell-Olds T, Dixon CJ, Kapralov MV, Filatov DA, Eyre-Walker A. 2010. Genome wide analyses reveal little evidence for adaptive evolution in many plant species. Mol Biol Evol. 27(8):1822–1832. doi:10.1093/molbev/msq079.

Hardigan MA, Laimbeer FPE, Newton L, Crisovan E, Hamilton JP, Vaillancourt B, Wiegert-Rininger K, Wood JC, Douches DS, Farré EM, *et al.* 2017. Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. Proc Natl Acad Sci U S A. 114(46):E9999–E10008. doi:10.1073/pnas.1714380114.

He L, Jia KH, Zhang RG, Wang Y, Shi TL, Li ZC, Zeng SW, Cai XJ, Wagner ND, Hörandl E, *et al.* 2021. Chromosome-scale assembly of the genome of *Salix dunnii* reveals a male-heterogametic sex determination system on chromosome 7. Mol Ecol Resour. 21(6):1966–1982. doi:10.1111/1755-0998.13362.

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet Res. 8(3):269–294. doi:10.1017/S0016672300010156.

Horvath DP, Patel S, Doğramaci M, Chao WS, Anderson JV, Foley ME, Scheffler B, Lazo G, Dorn K, Yan C, *et al.* 2018. Gene space and transcriptome assemblies of leafy spurge (*Euphorbia esula*) identify promoter sequences, repetitive elements, high-quality markers, and a full-length chloroplast genome. Weed Sci. 66(3):355–367. doi:10.1017/wsc.2018.2.

Hussin JG, Hodgkinson A, Idaghdour Y, Grenier J-C, Goulet J-P, Gbeha E, Hip-Ki E, Awadalla P. 2015. Recombination affects accumulation of damaging and disease-associated mutations in human populations. Nat Genet. 47(4):400–404. doi:10.1038/ng.3216.

Jalali S, Kancharla N, Yepuri V, Arockiasamy S. 2020. Exploitation of Hi-C sequencing for improvement of genome assembly and in-vitro validation of differentially expressing genes in *Jatropha curcas* L. 3 Biotech. 10(3):91. doi:10.1007/s13205-020-2082-0.

Jennings DL. 1963. Variation in pollen and ovule fertility in varieties of cassava, and the effect of interspecific crossing on fertility. Euphytica. 12(1):69–76. doi:10.1007/BF00033595.

Kantar MB, Nashoba AR, Anderson JE, Blackman BK, Rieseberg LH. 2017. The genetics and genomics of plant domestication. Bioscience. 67(11):971–982. doi:10.1093/biosci/bix114.

Katz K, Shutov O, Lapoint R, Kimelman M, Brister JR, O'Sullivan C. 2022. The sequence read archive: a decade more of explosive growth. Nucl Acids Res. 50(D1):D387–D390. https://doi.org/10.1093/nar/gkab1053.

Kocot D, Sitek E, Nowak B, Kołton A, Stachurska-Swakoń A, Towpasz K. 2022. The effectiveness of the sexual reproduction in selected clonal and nonclonal species of the genus *Ranunculus*. Biology (Basel). 11(1):85. doi:10.3390/biology11010085.

Kono TJY, Liu C, Vonderharr EE, Koenig D, Fay JC, Smith KP, Morrell PL. 2019. The fate of deleterious variants in a Barley genomic prediction population. Genetics. 213(4):1531–1544. doi:10.1534/genetics.119.302733.

Kryazhimskiy S, Bazykin GA, Dushoff J. 2008. Natural selection for nucleotide usage at synonymous and nonsynonymous sites in influenza A virus genes. J Virol. 82(10):4938–4945. doi:10.1128/JVI.02415-07.

Kubitzki K. 2014. Flowering plants. Eudicots: malpighiales. Berlin, Heidelberg: Springer-Verlag. p. 178911.

Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31(10):1674–1676. doi:10.1093/bioinformatics/btv033.

Lin X, Fang C, Liu B, Kong F. 2021. Natural variation and artificial selection of photoperiodic flowering genes and their applications in crop adaptation. aBIOTECH. 2(2):156–169. doi:10.1007/s42994-021-00039-0.

Liu J, Shi C, Shi CC, Li W, Zhang QJ, Zhang Y, Li K, Lu HF, Shi C, Zhu ST, *et al.* 2020. The chromosome-based rubber tree genome provides new insights into spurge genome evolution and rubber biosynthesis. Mol Plant. 13(2):336–350. doi:10.1016/j.molp.2019.10.017.

Liu H, Wei J, Yang T, Mu W, Song B, Yang T, Fu Y, Wang X, Hu G, Li W, *et al.* 2019. Molecular digitization of a botanical garden: high-depth whole-genome sequencing of 689 vascular plant species from the Ruili Botanical Garden. Gigascience. 8(4):giz007. doi:10.1093/gigascience/giz007.

Long EM, Romay MC, Ramstein G, Buckler ES, Robbins KR. 2023. Utilizing evolutionary conservation to detect deleterious mutations and improve genomic prediction in cassava. Front Plant Sci. 13:1041925. doi:10.3389/fpls.2022.1041925.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15(12):550. doi:10.1186/s13059-014-0550-8.

Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D, Goodstein DM, Schmutz J. 2022. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. ELife. 11:e78526. doi:10.7554/eLife.78526.

Lu J, Pan C, Fan W, Liu W, Zhao H, Li D, Wang S, Hu L, He B, Qian K, *et al.* 2022. A chromosome-level genome assembly of wild castor provides new insights into its adaptive evolution in tropical desert. Genom Proteom Bioinform. 20(1):42–59. doi:10.1016/j.gpb.2021.04.003.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 27(6):764–770. doi:10.1093/bioinformatics/btr011.

Marković V, Cvrčková F, Potocký M, Kulich I, Pejchar P, Kollárová E, Synek L, Žárský V. 2020. EXO70A2 is critical for exocyst complex

function in pollen development. Plant Physiol. 184(4):1823–1839. doi:10.1104/pp.19.01340.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature. 351(6328):652–654. doi:10.1038/351652a0.

Mckey D, Elias M, Pujol BB, Duputié A. 2010. Tansley review: the evolutionary ecology of clonally propagated domesticated plants. New Phytol. 186(2):318–332. doi:10.1111/j.1469-8137.2010.03210.x.

Messer PW, Petrov DA. 2013. Frequent adaptation and the McDonald-Kreitman test. Proc Natl Acad Sci U S A. 110(21):8615–8620. doi:10.1073/pnas.1220835110.

Michael TP, Salomé PA, Yu HJ, Spencer TR, Sharp EL, McPeek MA, Alonso JM, Ecker JR, McClung CR. 2003. Enhanced fitness conferred by naturally occurring variation in the circadian clock. Science. 302(5647):1049–1053. doi:10.1126/science.1082971.

Moyers BT, Morrell PL, Mckay JK. 2018. Genetic costs of domestication and improvement. J Hered. 109(2):103–116. doi:10.1093/jhered/esx069.

Muller HJ. 1964. The relation of recombination to mutational advance. Mutat Res. 1(1):2–9. doi:10.1016/0027-5107(64)90047-8.

Nakabayashi K, Okamoto M, Koshiba T, Kamiya Y, Nambara E. 2005. Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: epigenetic and genetic regulation of transcription in seed. Plant J. 41(5):697–709. doi:10.1111/j.1365-313X.2005.02337.x.

Ng PC, Henikoff S. 2003. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 31(13):3812–3814. doi:10.1093/nar/gkg509.

Ogbonna AC, Ramu P, Esuma W, Nandudu L, Morales N, Powell A, Kawuki R, Bauchet G, Jannink JL, Mueller LA. 2021. A population based expression atlas provides insights into disease resistance and other physiological traits in cassava (*Manihot esculenta* Crantz). Sci Rep. 11(1):23520. doi:10.1038/s41598-021-02794-y.

Oluwasanya D, Esan O, Hyde PT, Kulakow P, Setter TL. 2021. Flower development in cassava is feminized by cytokinin, while proliferation is stimulated by anti-ethylene and pruning: transcriptome responses. Front Plant Sci. 12:666266. doi:10.3389/fpls.2021.666266.

Parmar A, Sturm B, Hensel O. 2017. Crops that feed the world: production and improvement of cassava for food, feed, and industrial uses. Food Secur. 9(5):907–927. doi:10.1007/s12571-017-0717-8.

Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. 2013. Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet. 9(8):e1003709. doi:10.1371/journal.pgen.1003709.

Pineda M, Yu B, Tian Y, Morante N, Salazar S, Hyde PT, Setter TL, Ceballos H. 2020. Effect of pruning young branches on fruit and seed set in cassava. Front Plant Sci. 11:1107. doi:10.3389/fpls.2020.01107.

Pootakham W, Sonthirod C, Naktang C, Ruang-Areerate P, Yoocha T, Sangsrakru D, Theerawattanasuk K, Rattanawong R, Lekawipat N, Tangphatsornruang S. 2017. De novo hybrid assembly of the rubber tree genome reveals evidence of paleotetraploidy in Hevea species. Sci Rep. 7(1):41457. doi:10.1038/srep41457.

Pujol B, David P, McKey D. 2005. Microevolution in agricultural environments: how a traditional Amerindian farming practice favours heterozygosity in cassava (*Manihot esculenta* Crantz, Euphorbiaceae). Ecol Lett. 8(2):138–147. doi:10.1111/j.1461-0248.2004.00708.x.

Qi W, Lim YW, Patrignani A, Schläpfer P, Bratus-Neuenschwander A, Grüter S, Chanez C, Rodde N, Prat E, Vautrin S. 2022. The haplotype-resolved chromosome pairs of a heterozygous diploid African cassava cultivar reveal novel pan-genome and allele-specific transcriptome features. Gigascience. 11:giac028. doi:10.1093/gigascience/giac028.

Qin P, Tu B, Wang Y, Deng L, Quilichini TD, Li T, Wang H, Ma B, Li S. 2013. ABCG15 encodes an ABC transporter protein, and is essential for post-meiotic anther and pollen exine development in rice. Plant Cell Physiol. 54(1):138–154. doi:10.1093/pcp/pcs162.

Ramos Abril LN, Pineda LM, Wasek I, Wedzony M, Ceballos H. 2019. Reproductive biology in cassava: stigma receptivity and pollen tube growth. Commun Integr Biol. 12(1):96–111. doi:10.1080/19420889.2019.1631110.

Ramu P, Esuma W, Kawuki R, Rabbi IY, Egesi C, Bredeson JV, Bart RS, Verma J, Buckler ES, Lu F. 2017. Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. Nat Genet. 49(6):959–963. doi:10.1038/ng.3845.

Rodgers-Melnick E, Bradbury PJ, Elshire RJ, Glaubitz JC, Acharya CB, Mitchell SE, Li C, Li Y, Buckler ES, Rodgers-Melnick E, *et al.* 2015. Recombination in diverse maize is stable, predictable, and associated with genetic load. Proc Natl Acad Sci U S A. 112(12):3823–3828. doi:10.1073/pnas.1413864112.

Rojas MC, Pérez JC, Ceballos H, Baena D, Morante N, Calle F. 2009. Analysis of inbreeding depression in eight S_1 cassava families. Crop Sci. 49(2):543–548. doi:10.2135/cropsci2008.07.0419.

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Schölkopf B, Weigel D, Lohmann JU. 2005. A gene expression map of *Arabidopsis thaliana* development. Nat Genet. 37(5):501–506. doi:10.1038/ng1543.

Schulz AJ, Zhai J, AuBuchon-Elder T, El-Walid M, Ferebee TH, Gilmore EH, Hufford MB, Johnson LC, Kellogg EA, La T, *et al.* 2023. Fishing for a reelGene: evaluating gene models with evolution and machine learning. BioRxiv 558246. https://doi.org/10.1101/2023.09.19.558246, preprint: not peer reviewed.

Silva DD, Martins ML, Santos AS, Santos VD, Alves AA, Ledo CA. 2018. Obtaining hybrids of cultivars and wild subspecies of cassava. Pesqui Agropecu Bras. 53(2):182–188. doi:10.1590/s0100-204x2018000200006.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31(19):3210–3212. doi:10.1093/bioinformatics/btv351.

Simmonds NW. 1997. A review of potato propagation by means of seed, as distinct from clonal propagation by tubers. Potato Res. 40:191–214. https://doi.org/10.1007/BF02358245.

Souza LS, Alves AA, de Oliveira EJ. 2020. Phenological diversity of flowering and fruiting in cassava germplasm. Sci Hortic. 265:109253. doi:10.1016/j.scienta.2020.109253.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30(9):1312–1313. doi:10.1093/bioinformatics/btu033.

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam M, Ralph S, Rombauts S, Salamov A, *et al.* 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science. 313(5793):1596–1604. doi:10.1126/science.1128691.

Ukwu NU, Olasanmi B. 2018. Crossability among five cassava (*Manihot esculenta* CRANTZ) varieties. Mod Concepts Dev Agron. 2(4):1–6. doi:10.31031/MCDA.2018.02.000543.

Vieira LD, Soares TL, Rossi ML, Alves AA, Santos FD, Souza FV. 2012. Viability, production and morphology of pollen grains for different species in the genus Manihot (Euphorbiaceae). Acta Bot Bras. 26(2):350–356. doi:10.1590/S0102-33062012000200011.

Waese J, Fan J, Pasha A, Yu H, Fucile G, Shi R, Cumming M, Kelley LA, Sternberg MJ, Krishnakumar V, *et al.* 2017. Eplant: visualizing and exploring multiple levels of data for hypothesis generation in

plant biology. Plant Cell. 29(8):1806–1821. doi:10.1105/tpc.17.00073.

Wang W, Feng B, Xiao J, Xia Z, Zhou X, Li P, Zhang W, Wang Y, Møller BL, Zhang P, *et al.* 2014. Cassava genome from a wild ancestor to cultivated varieties. Nat Commun. 5(1):5110. doi:10.1038/ncomms6110.

Wei S, Yang Y, Yin T. 2020. The chromosome-scale assembly of the willow genome provides insight into Salicaceae genome evolution. Hortic Res. 7(1):45. doi:10.1038/s41438-020-0268-6.

Wiens D, Calvin CL, Wilson CA, Davern CI, Frank D, Seavey SR. 1987. Reproductive success, spontaneous embryo abortion, and genetic load in flowering plants. Oecologia. 71(4):501–509. doi:10.1007/BF00379288.

Wu H, Yao D, Chen Y, Yang W, Zhao W, Gao H, Tong C. 2020. De novo genome assembly of populus simonii further supports that populus simonii and populus trichocarpa belong to different sections. G3 (Bethesda). 10(2):455–466. doi:10.1534/g3.119.400913.

Wurdack KJ, Hoffmann P, Chase MW. 2005. Molecular phylogenetic analysis of uniovulate Euphorbiaceae (*Euphorbiaceae sensu stricto*) using plastid RBCL and TRNL-F DNA sequences. Am J Bot. 92(8):1397–1420. doi:10.3732/ajb.92.8.1397.

Xu S, He Z, Zhang Z, Guo Z, Guo W, Lyu H, Li J, Yang M, Du Z, Huang Y, *et al.* 2017. The origin, diversification and adaptation of a major mangrove clade (Rhizophoreae) revealed by whole-genome sequencing. Natl Sci Rev. 4(5):721–734. doi:10.1093/nsr/nwx065.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591. doi:10.1093/molbev/msm088.

Zhang M, Zhou L, Bawa R, Suren H, Holliday JA. 2016. Recombination rate variation, hitchhiking, and demographic history shape deleterious load in poplar. Mol Biol Evol. 33(11):2899–2910. doi:10.1093/molbev/msw169.

Zhang R, Murat F, Pont C, Langin T, Salse J. 2014. Paleo-evolutionary plasticity of plant disease resistance genes. BMC Genomics. 15(1):187. doi:10.1186/1471-2164-15-187.

Zhou R, Macaya-Sanz D, Carlson CH, Schmutz J, Jenkins JW, KudrnaD, Sharma A, Sandor L, Shu S, Barry K, *et al.* 2020. A willow sex chromosome reveals convergent evolution of complex palindromic repeats. Genome Biol. 21(1):38. doi:10.1186/s13059-020-1952-4.

Zhou W, Wang Y, Li B, Petijová L, Hu S, Zhang Q, Niu J, Wang D, Wang S, Dong Y, *et al.* 2021. Whole-genome sequence data of *Hypericum perforatum* and functional characterization of melatonin biosynthesis by N-acetylserotonin O-methyltransferase. J Pineal Res. 70(2):e12709. doi:10.1111/jpi.12709.

Zhou Y, Massonnet M, Sanjak JS, Cantu D, Gaut BS. 2017. Evolutionary genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. Proc Natl Acad Sci U S A. 114(44):11715–11720. doi:10.1073/pnas.1709257114.

Editor: A. Kern