Studies on the hybrid origin of Guinea yam and its evolution

Yu Sugihara

TABLE OF CONTENTS

ABSTRACT

CHAPTER 1: GENEARL INTRODUCTION

Page 4: THE GENUS *DIOSCOREA*

Page 9: DOMESTICATION OF YAM

CHAPTER 2: HYBRID ORIGIN OF GUINEA YAM AS REVEALED BY GENOME ANALYSIS

Page 10: INTRODUCTION

Page 12: RESULTS

Page 12: GENETIC DIVERSITY OF GUINEA YAM

Page 17: PHYLOGENOMIC ANALYSIS OF AFRICAN YAM

Page 20: HYBRID ORIGIN OF GUINEA YAM

Page 25: EVOLUTIONARY HISTORY OF GUINEA YAM

Page 28: EXTENSIVE INTROGRESSION AT THE *SWEETIE* LOCUS

Page 30: DISCUSSION

Page 30: HOMOPLOID HYBRIDIZATION AS THE TRIGGER OF DOMESTICATION

Page 30: USE OF WILD SPECIES TO IMPROVE GUINEA YAM

Page 32: MATERIALS AND METHODS

CHAPTER 3: GENEARL DISCUSSION

SUPPLEMENTARY DATA

ACKNOWLEDGEMENTS

REFERENCES

ABSTRACT

Yam is a collective name of tuber crops belonging to the genus *Dioscorea*. Yam is important not only as a staple food crop but also as an integral component of society and culture of the millions of people who depend on it. However, due to its regional importance, yam has long been regarded as an 'orphan crop' lacking a due global attention. Although this perception is changing with recent advances in genomics technologies, domestication processes of most yam species are still ambiguous. This is mainly due to the complicated evolutionary history of *Dioscorea* species caused by frequent hybridization and polyploidization, which is possibly caused by dioecy that imposed obligate outcrossing to the species of *Dioscorea*. White Guinea yam (*Dioscorea rotundata*) is an important staple tuber crop in West Africa. However, its origin remains unclear. In this study, we resequenced 336 accessions of white Guinea yam and compared them with the sequences of wild *Dioscorea* species using an improved reference genome sequence of *D. rotundata*. In contrast to a previous study suggesting that *D. rotundata* originated from a subgroup of *Dioscorea praehensilis*, our results suggest a hybrid origin of white Guinea yam from crosses between the wild rainforest species *D. praehensilis* and the savannah-adapted species *D. abyssinica*. We identified a greater genomic contribution from *D. abyssinica* in the sex chromosome of Guinea yam. The haplotype network of the chroloplast sequences of both diploid and triploid *D. rotundata* and its wild relatives showed that the female parent of *D. rotundata* was *D. abyssinica* and the male parent was *D. praehensilis*. We also found extensive introgression around the *SWEETIE* gene. Our findings point to a complex domestication scenario for Guinea yam and highlight the importance of wild species as gene donors for improving this crop through molecular breeding.

CHAPTER 1: GENERAL INTRODUCTION

THE GENUS *DIOSCOREA*

The genus *Dioscorea*, which consists of approximately 630 species, is the largest one in the family Dioscreaceae of monocotyledons (WCSP, 2020). It is widely distributed in the tropical and temperate regions and occurs in diverse environments from forests to grasslands (Maurin et al., 2016; Viruel et al., 2016; Wilkin et al., 2005). Several studies have been conducted on the phylogenetic relationships of species in *Dioscorea*. Previously, intrageneric taxa have been proposed based on morphological characters (Burkill, 1960). However, diagnostic keys and delineation of taxa varied according to the authors. Recently, phylogenetic analyses have been conducted based on chloroplast DNA (cpDNA) sequences and nuclear gene sequences (Noda et al., 2020). Noda et al. (2020) provided a large-scale phylogenetic tree containing 183 species and proposed dividing *Dioscorea* into two subgenera (*Dioscorea* and *Helmia*), with 11 major clades and 27 sections/species groups.

Dioscorea likely originated in the Laurasian Palaearctic between the Late Cretaceous and the Early Eocene (Fig. 1.1). In the Eocene and Oligocene, *Dioscorea* expanded to the southern region by long-distance dispersal or migration by land bridges. In the Oligocene and Miocene, main *Dioscorea* lineages experienced divergence events on a world-wide scale. In the Miocene and Pliocene, some lineages dispersed into new areas. The number of biogeographical speciation events seems to have decreased after the Quaternary period began (Couto et al., 2018; Maurin et al., 2016; Viruel et al., 2016).

The majority of *Dioscorea* species are perennial herbaceous climbers with simple or compound leaves and reproduce sexually and/or clonally (Fig. 1.2). Flowers in *Dioscorea* are mostly dioecious with male and female flowers borne on separate individuals, and multiple sex-determination systems (XY or ZW) were reported in the genus (Cormier et al., 2019; Tamiru et al., 2017; Terauchi & Kahl, 1999). Most species produce winged seeds and capsular, six-seeded fruits, while some species have wingless seeds, samaroid or berry fruits (Caddick et al., 2002; Noda et al., 2020). In addition to sexual reproduction, *Dioscorea* species propagate clonally by bulbils, rhizomes or tubers. Bulbils are aerial tubers that are formed in the axils of leaves or bracts

of some *Dioscorea* species (Fig. 1.2f). They are mainly consumed as food, but also used as folk medicine in many cultures (Ikiriza et al., 2019). Bulbils are generally brown-colored and have small tubercles over their surface, but their shape and size vary in the different species (Murty & Purnima, 1983). *D. bulbifera* (also known as aerial yam) is the major bulbil-producing species and is characterized by considerable bulbil shape diversity (Terauchi et al., 1991). Rhizomes and tubers represent morphologically diverse structures that serve as underground starch storage organs (Fig. 1.3). Because these storage organs serve as food sources for various wild animals, they have evolved defense traits. For example, *D. praehensilis* has crown roots with spines to protect tubers from burrowing or digging animals (Fig. 1.3c). Some species of the African clade have thick corky barks covering the pachycaul structure that may provide protection against fire and herbivores (Maurin et al., 2016). In addition, *Dioscorea* species produce diverse secondary metabolites such as saponins, alkaloids, and tannins that serve a variety of functions including defense against herbivores (Coursey, 1967). Chemical components of some species have medicinal values (Dutta, 2015; Liu et al., 2008).

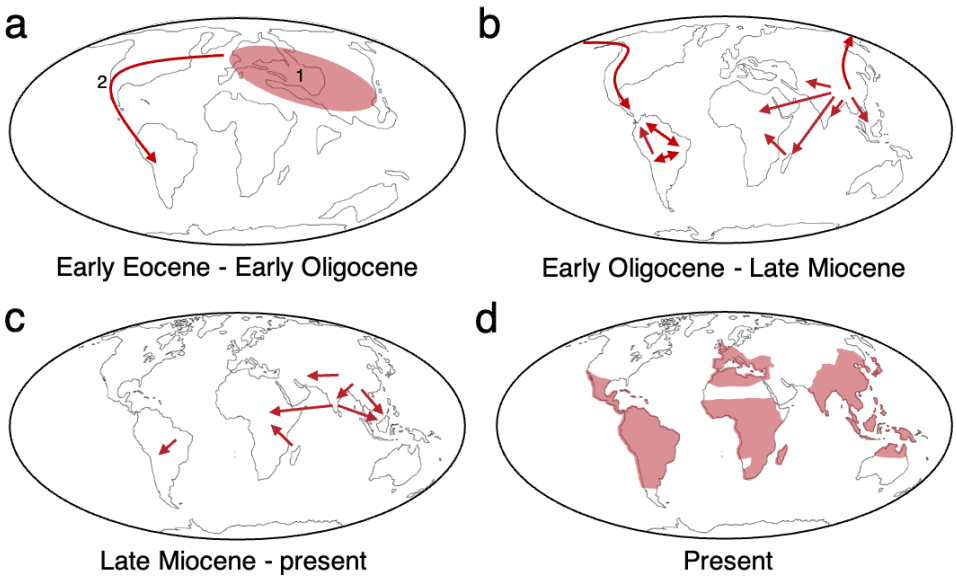


Fig. 1.1 Biogeographical origin and distribution of *Dioscorea* species (Viruel et al., 2016). **a** *Dioscorea* likely originated in the Laurasian Palaearctic in the Late Cretaceous and the Early Eocene (1) and then dispersed from Asia to South America (2). **b** In the Oligocene and Miocene, *Dioscorea* mainly expanded to the southern region. **c** Some lineages dispersed into new areas in the Miocene and Pliocene, but speciation events decreased in the Quaternary. **d** Geographical distribution in the present era. (Maps are based on C. R. Scotese's PALEOMAP project; www.scotese.com).

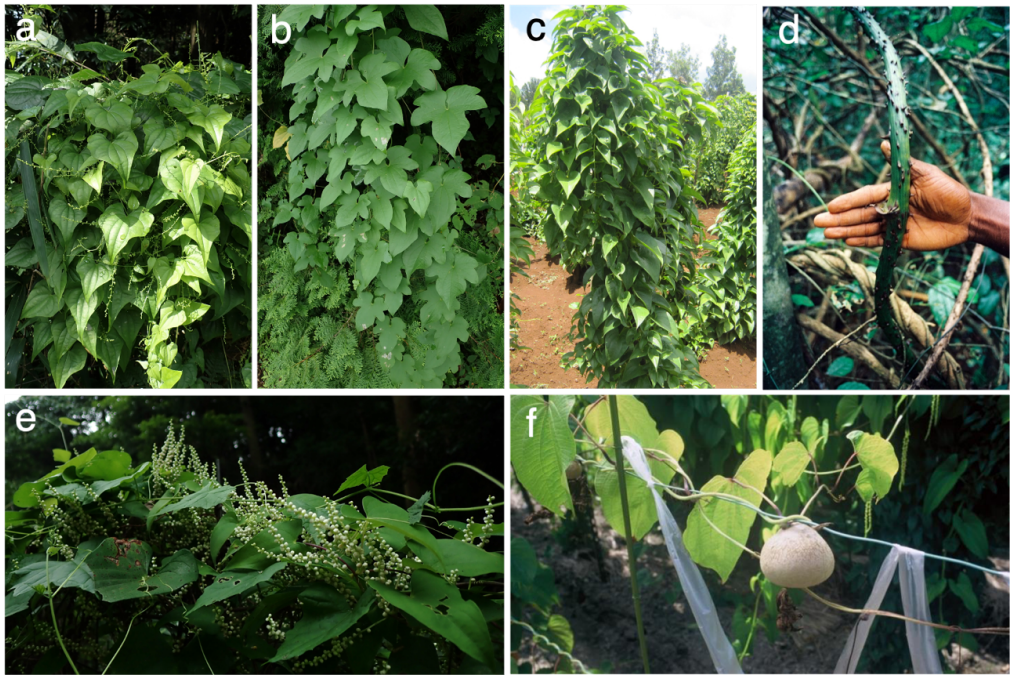


Fig. 1.2 Morphological diversity of the above-ground parts of *Dioscorea* species **a** *D. tokoro*, **b** *D. quinqueloba*, **c** *D. rotundata*, **d** a stem of *D. mangenotiana* with thorns, **e** flowers of *D. japonica*, **f** a bulbil of *D. bulbifera*.

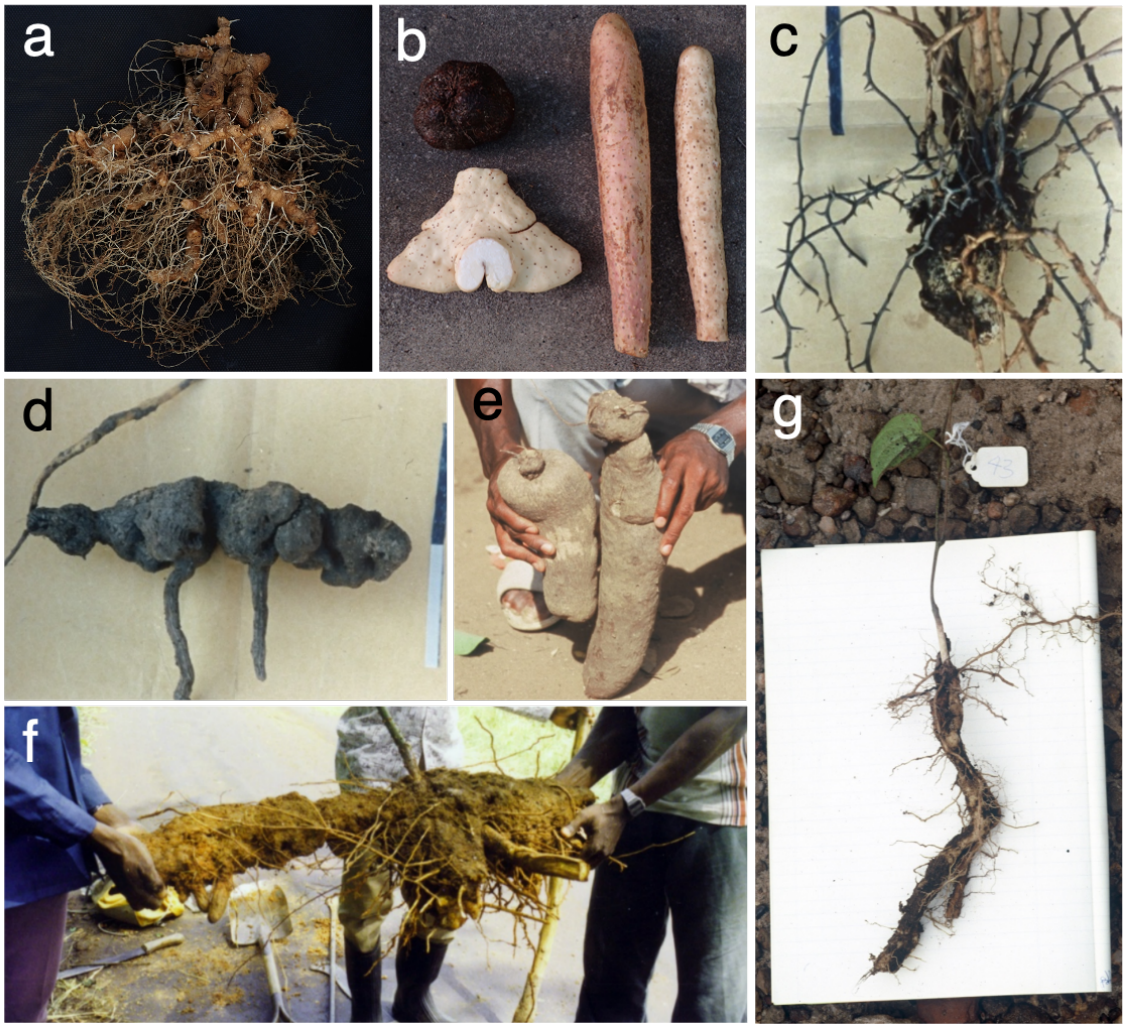


Fig. 1.3 Rhizomes and tubers of *Dioscorea* species. **a** rhizomes of *D. tokoro*, **b** Tukuneimo group (left top), Ichoimo group (left bottom), Nagaimo group (right) in *D. polystachya*, **c** *D. praehensilis*, **d** *D. minutiflora*, **e** *D. rotundata* (left), *D. cayenensis* (right), **f** *D. mangenotiana*, **g** *D. abyssinica*.

DOMESTICATION OF YAM

Yam is a collective name of tuber crops belonging to the genus *Dioscorea*. In 2018, the global yam production was around 72.6 million tons (FAOSTAT, 2018). The major yam species include *Dioscorea rotundata*, *D. alata*, *D. trifida*, *D. polystachya*, and *D. esculenta* (Arnau et al., 2010). White Guinea yam (*D. rotundata*) is the most important yam worldwide, accounting for ~92.5% of the total world yam production (FAOSTAT, 2018). Guinea yam is mainly grown in West and Central Africa, especially in Côte d'Ivoire, Ghana, Togo, Benin, Nigeria, and Cameroon, the region known as the 'yam belt'. By contrast, greater yam (*D. alata*) that originated in Asia is the most widely distributed species in the world. Yam is a staple crop in many tropical countries, and it also plays important roles in society and culture of the people in the major yam-growing regions (Coursey, 1972; Obidiegwu et al., 2020; Obidiegwu & Akpabio, 2017). However, due to its localized importance, yam has been regarded as an 'orphan crop' and received considerably less research attention compared to the major crop species.

Yams of different *Dioscorea* species are believed to be independently domesticated in different continents: *D. rotundata* and *D. cayenensis* in West and Central Africa, *D. alata* in Southeast Asia, and *D. trifida* in South America. However, our knowledge of their origins has been limited until recently. This is mainly due to the frequent hybridization and polyploidization of many species including *D. rotundata* (Chaïr et al., 2010; Girma et al., 2014; Scarcelli et al., 2006, 2017; Sugihara et al., 2020; Terauchi et al., 1992) and *D. alata* (Chaïr et al., 2016; Sharif et al., 2020). The recent population genomics studies have started unveiling the domestication processes of the major species (Scarcelli et al., 2019; Sharif et al., 2020; Sugihara et al., 2020).

CHAPTER 2: HYBRID ORIGIN OF GUINEA YAM AS REVEALED BY GENOME ANALYSIS

INTRODUCTION

Yams (*Dioscorea* spp.) are major starchy tuber crops that are widely consumed in the tropics. Ten yam species are cultivated worldwide, including *D. alata* in Southeast Asia, *D. trifida* in South America, and *D. rotundata* in West and Central Africa (Hancock, 2012). *D. rotundata*, also known as white Guinea yam, is the most important species in West and Central Africa, an area that accounted for 92.5% of global yam production in 2018 (FAOSTAT, 2018). Beyond its nutritional and food value, Guinea yam is also important for the culture of West African people (Obidiegwu & Akpabio, 2017).

Despite the considerable importance of Guinea yam, its origin has been elusive. There are two types of Guinea yam: white Guinea yam (*D. rotundata*) and yellow Guinea yam (*D. cayenensis*). *D. cayenensis* is thought to be a triploid species of hybrid origin, with *D. rotundata* as the maternal parent and *D. burkilliana* as the paternal parent (Girma et al., 2014; Terauchi et al., 1992). In turn, the triploid *D. rotundata* is thought to be a hybrid between *D. rotundata* and *D. togoensis* (Girma et al., 2014). However, the origin of diploid *D. rotundata*, which represents the majority of Guinea yam (Girma et al., 2014), has been ambiguous. Two wild species are candidate progenitors of diploid *D. rotundata*: the savannah-adapted wild species *D. abyssinica* and the rainforest-adapted wild species *D. praehensilis* (Coursey, 1976a, 1976b; Girma et al., 2014; Magwé-Tindo et al., 2018; Scarcelli et al., 2006, 2017, 2019; Terauchi et al., 1992). The geographical distributions of *D. abyssinica* and *D. praehensilis* slightly overlap (Fig. 2.1). Based on morphological evaluation, D.G. Coursey (1976) proposed that *D. rotundata* might be a hybrid between the two species (Coursey, 1976b). However, other reports indicate that the origin of Guinea yam is ambiguous due to the small number of markers (Girma et al., 2014; Magwé-Tindo et al., 2018; Scarcelli et al., 2006, 2017; Terauchi et al., 1992) or introgression (Scarcelli et al., 2006, 2017) or incomplete lineage sorting (Scarcelli et al., 2017).

The whole-genome sequence of Guinea yam has been reported (Tamiru et al., 2017). A recent genome study involving 86 *D. rotundata*, 47 *D. praehensilis*, and 34 *D. abyssinica* accessions suggested that diploid *D. rotundata* was domesticated from *D. praehensilis* (Scarcelli et al., 2019). Here, we addressed this hypothesis using an expanded set of genomes from cultivated and wild *Dioscorea* species.

In this study, we generated an improved version of the Guinea yam reference genome and used it to analyze the genomes of 336 accessions of *D. rotundata* and its wild relatives. Based on these analyses, we attempted to reveal the history of Guinea yam domestication. Our results suggest that diploid *D. rotundata* was most likely derived from homoploid hybridization between *D. abyssinica* and *D. praehensilis*. By evaluating the genomic contributions of each parental species to *D. rotundata*, we revealed higher representation of the *D. abyssinica* genome in the sex chromosome of *D. rotundata* and a signature of extensive introgression in the *SWEETIE* gene on chromosome 17.

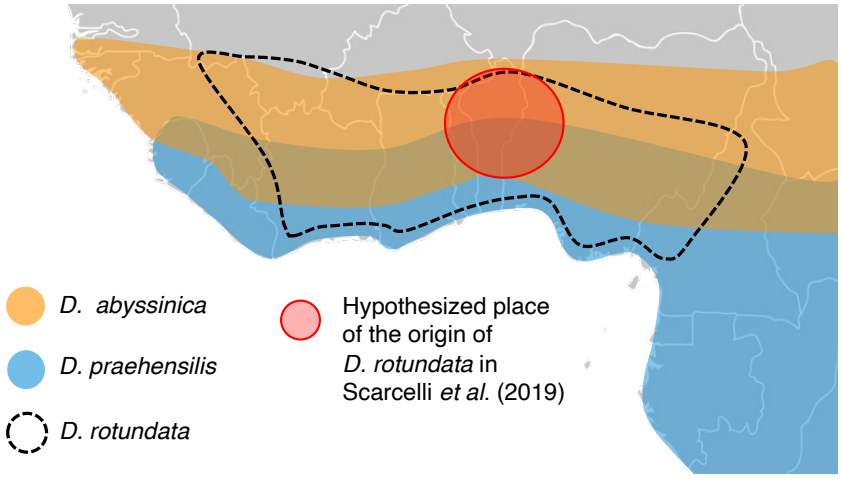


Fig. 2.1. The geographical distributions of African yams. Adapted from Scarcelli *et al*. (2017) and Scarcelli *et al*. (2019).

RESULTS

Genetic diversity of Guinea yam

We obtained DNA samples from 336 accessions of *D. rotundata* maintained at the International Institute of Tropical Agriculture (IITA), Nigeria, representing the genetic diversity of Guinea yam landraces and improved lines from West Africa. We subjected these samples to whole-genome resequencing on the Illumina sequencing platform. We aligned the resulting short reads to the newly assembled reference genome (material and method S1 and S2) and extracted SNP information for use in genetic diversity studies (Table S1, S2, and material and method S3). Based on admixture analysis by sNMF (Frichot et al., 2014), we defined five major clusters (Fig. 2.2A). When K = 2, cluster 1 was clearly separated from the other accessions. Principal component analysis (PCA) also separated cluster 1 from the rest of the clusters (Fig. 2.2B). Accessions in cluster 1 had significantly higher heterozygosity and ~10-times more unique alleles than those in the four remaining clusters (Fig. 2.3–2.4, and Table 2.1). Because flow cytometry analysis confirmed that all 10 accessions analyzed in cluster 1 were triploids (Table S1), we hypothesized that cluster 1 represents triploid *D. rotundata*, which was reported to be a hybrid between *D. rotundata* and *D. togoensis* (Girma et al., 2014). After removing the cluster 1 accessions, the nucleotide diversity of *D. rotundata* was estimated to be 14.83 x 10^{-4} (Table 2.2) which is approximately 1.5 times larger than that reported previously (Scarcelli et al., 2019), presumably because we used a larger number of samples with diverse genetic backgrounds in our study. Linkage disequilibrium (LD) of diploid *D. rotundata* showed a decay of r^2 = 0.13 in a 200-kb genomic region (Fig. 2.5), which is slower than that of cassava, another clonally propagated crop (Ramu et al., 2017).

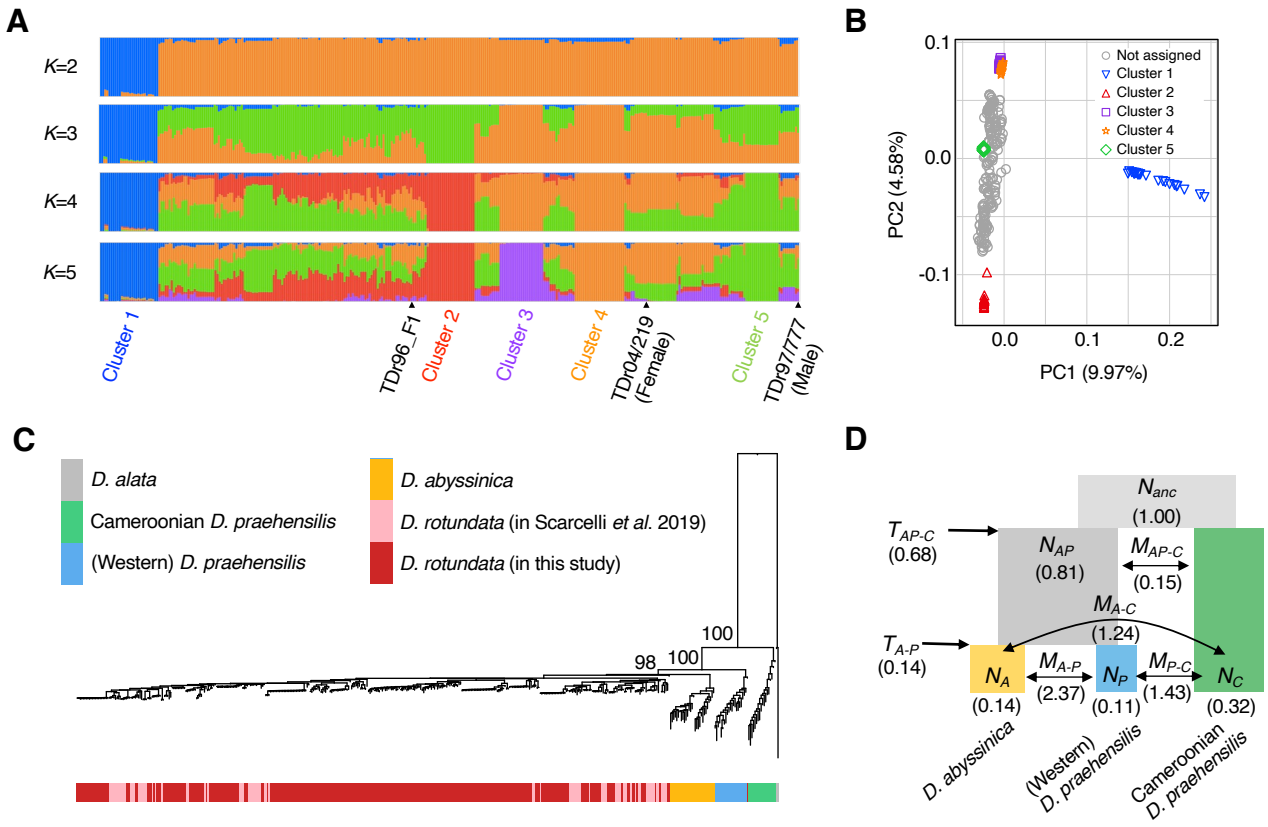


Fig. 2.2. Genetic diversity and phylogenomics of Guinea yam and its wild relatives. (A) Ancestry proportions of each Guinea yam accession with 6,124,093 SNPs. "TDr96_F1" is the sample used as the reference genome. **(B)** PCA result of the 336 Guinea yam accessions. **(C)** Neighbor-joining tree of four African yam lineages reconstructed using *D. alata* as an outgroup based on 463,293 SNPs. The numbers indicate bootstrap values after 100 replications. The sequences of *D. rotundata* in the previous study (Scarcelli et al., 2019) were included in the tree. **(D)** Evolutionary relationship of three African wild yam lineages (*D. abyssinica*, Western *D. praehensilis*, Cameroonian *D. praehensilis*) as inferred by ∂a∂i (Gutenkunst et al., 2009) using 17,532 SNPs. *N*, *M*, and *T* represent the relative population size from N_{anc}, migration rate, and divergence time, respectively.

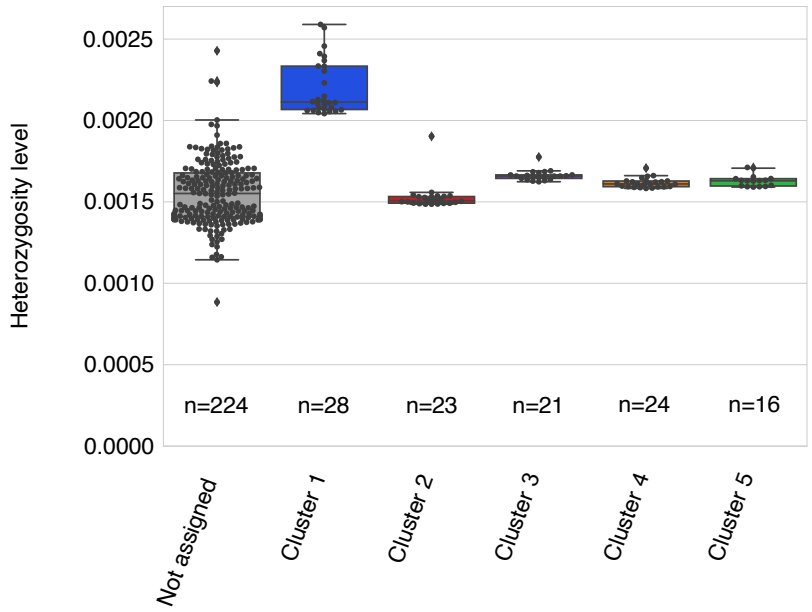


Fig. 2.3. Heterozygosity levels of samples in five clusters of *D. rotundata*. Heterozygosity level of an individual is defined as the ratio of number of heterozygous SNPs to the total number of mapped sites to the reference genome.

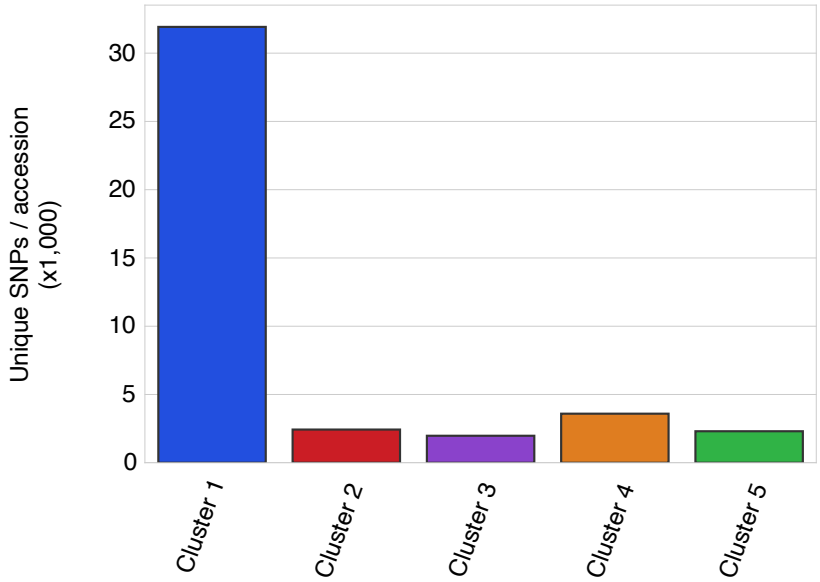


Fig. 2.4. Number of unique alleles in the five clusters of *D. rotundata*.

Table 2.1. Comparison of heterozygosity levels in the five clusters of *D. rotundata*. Heterozygosity level of an individual is defined as the ratio of number of heterozygous SNPs to the total number of mapped sites to the reference genome. The diagonal cell represents the mean ± standard deviation of the heterozygosity levels of the samples in each cluster. The other cells represent *P*-values of the difference of the heterozygosity levels between the two clusters as calculated by two-tailed Student t test. Cluster 1 has a significantly higher heterozygosity level than the other clusters.

	Not assigned					
Not assigned	15.53×10^{-4} ($\pm1.96\times10^{-4}$)	**Cluster 1**				
Cluster 1 (n=28)	2.874×10^{-42}	21.98×10^{-4} ($\pm1.68\times10^{-4}$)	**Cluster 2**			
Cluster 2 (n=23)	0.5483	1.453×10^{-22}	15.29×10^{-4} ($\pm0.84\times10^{-4}$)	**Cluster 3**		
Cluster 3 (n=21)	0.01194	8.305×10^{-19}	2.582×10^{-8}	16.62×10^{-4} ($\pm0.32\times10^{-4}$)	**Cluster 4**	
Cluster 4 (n=24)	0.1188	4.358×10^{-22}	1.759×10^{-5}	9.915×10^{-6}	16.16×10^{-4} ($\pm0.30\times10^{-4}$)	**Cluster 5**
Cluster 5 (n=16)	0.1203	1.344×10^{-16}	6.272×10^{-5}	7.857×10^{-3}	0.1972	16.30×10^{-4} ($\pm0.37\times10^{-4}$)

Table 2.2. Population genetics summary statistic in the 308 yam accessions

	After imputation
No. segregating site	5,229,368
No. singleton	1,227,900
θ_W	14.98 x 10^{-4}
θ_π	14.83 x 10^{-4}
Tajima's *D*	-0.0305

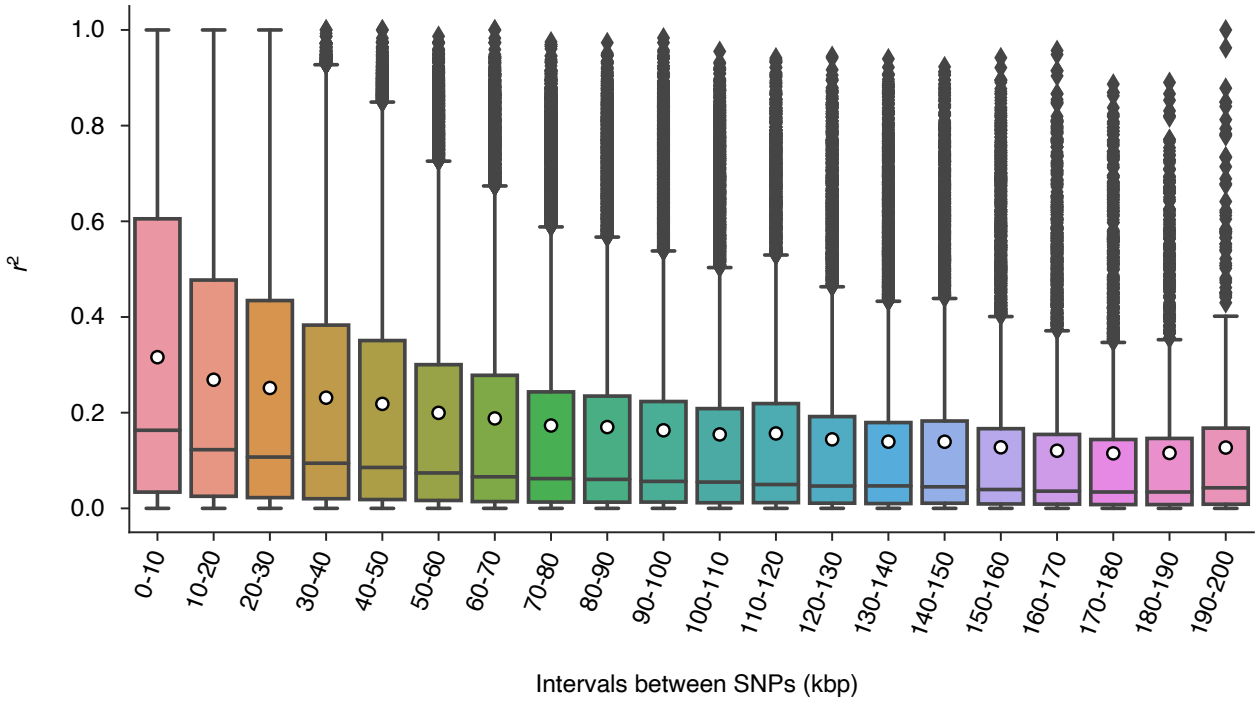


Fig. 2.5. LD decay of *D. rotundata*. Each white dot represents the average r^2 in each interval.

Phylogenomic analysis of African yam

Using the SNP information, we constructed a rooted neighbor-joining (NJ) tree (Saitou & Nei, 1987) based on 308 Guinea yam accessions sequenced in the present study (excluding cluster 1 triploid accessions), as well as 80 *D. rotundata*, 29 *D. abyssinica*, 21 Western *D. praehensilis*, and 18 Cameroonian *D. praehensilis* accessions that were sequenced in a previous study (Scarcelli et al., 2019) using two accessions of Asian species *D. alata* as an outgroup (Fig. 2.2C). Throughout the analyses described below, we used 388 *D. rotundata* accessions by combining our samples and those used previously (Scarcelli et al., 2019). According to this NJ tree, the *D. rotundata* accessions sequenced in this study are genetically close to the *D. rotundata* accessions reported previously (Scarcelli et al., 2019) (Fig. 2.2C). However, the NJ tree showed that *D. rotundata* is more closely related to *D. abyssinica* than to Western *D. praehensilis* (Fig. 2.2C), which is inconsistent with the previous finding (Scarcelli et al., 2019) that *D. rotundata* is most closely related to Western *D. praehensilis*.

To elucidate the evolutionary relationships of the three wild *Dioscorea* species that are closely related to *D. rotundata*, *D. abyssinica* (indicated as A), Western *D. praehensilis* (P), and Cameroonian *D. praehensilis* (C), we performed Diffusion Approximations for Demographic Inference (∂a∂i) analysis (Gutenkunst et al., 2009), which allows demographic parameters to be estimated based on an unfolded site frequency spectrum. First, we tested three phylogenetic models, {{A, P}, C}, {{P, C}, A}, and {{C, A}, P}, using 17,532 SNPs that were polarized using *D. alata* as an outgroup without considering migration among the species. Of the three models, {{A, P}, C} had the highest likelihood (Table 2.3).

This result is not consistent with the finding (Scarcelli et al., 2019) that {{P, C}, A} had the highest likelihood, as determined using a different method with fastsimcoal2 software (Excoffier et al., 2013). To exactly repeat the previous analysis, we tested these three models with fastsimcoal2 (Excoffier et al., 2013) using the previous reference genome (Tamiru et al., 2017), which indicated that {{A, P}, C} had the highest likelihood (Table 2.4). Taken together, our results are not consistent with the previous report (Scarcelli et al., 2019). However, they are consistent with the PCA result from the same report, which separated Cameroonian *D. praehensilis* from the other African yams in PC1 (Fig. 2A of Scarcelli et al., 2019).

Based on the assumption that {{A, P}, C} describes the true evolutionary relationship among the three wild *Dioscorea* species, we re-estimated the evolutionary parameters with ∂a∂i, allowing symmetric migration (gene flow) among the species (Fig. 2.2D). Since the results indicated that Cameroonian *D. praehensilis* is distantly related to *D. rotundata* and was not likely involved in genetic exchange with *D. rotundata* (Fig. 2.2C), we focused on Western *D. praehensilis*, which we will refer to as *D. praehensilis* for brevity.

Table 2.3. Likelihood comparison in ∂a∂i

Model	$\log_{10}(L)$	No. parameters	AIC	Illustration of the model
{{A, P}, C} (without migration)	-15289.70	6	30591.40	-
{{P, C}, A} (without migration)	-15765.32	6	31542.64	-
{{C, A}, P} (without migration)	-15765.15	6	31542.29	-
{{A, P}, C} (with migration)	-12739.86	10	25499.72	Fig. 2.2D
{{A, R}, P} (with migration)	-10149.73	10	20319.47	-
{{P, R}, A} (with migration)	-10385.46	10	20790.92	-
{{A, R}, {P, R}} (with migration)	-10052.96	9	20123.91	Fig. 2.6C
{{A, R}, {P, R}} - With migration - With population growth - Fix the parameters except for population size	-10046.73	6	20105.47	Fig. 2.8C

C: Cameroonian *D. praehensilis*

A: *D. abyssinica*

P: (Western) *D. praehensilis*

R: *D. rotundata*

Table 2.4. Likelihood comparison in fastsimcoal2

Model	$\log_{10}(L)$
{{A, P}, C} (without migration)	-172110.065
{{P, C}, A} (without migration)	-174281.072
{{C, A}, P} (without migration)	-173358.592

Hybrid origin of Guinea yam

We propose three hypotheses for the origin of Guinea yam (*D. rotundata*) based on the NJ tree (Fig. 2.2C) and ∂a∂i (Gutenkunst et al., 2009) (Fig. 2.2D). The first hypothesis is that *D. rotundata* was derived from *D. abyssinica* (Hypothesis 1 in Fig. 2.6A). The second is that *D. rotundata* was derived from *D. praehensilis* (Hypothesis 2 in Fig. 2.6A). However, in Hypotheses 1 and 2, the divergence time of *D. rotundata* from the wild species may not be sufficient to separate the three lineages, and there may be incomplete lineage sorting among the species. The third hypothesis is that *D. rotundata* originated as an admixture between *D. abyssinica* and *D. praehensilis* (Hypothesis 3 in Fig. 2.6A).

Before estimating the evolutionary parameters for the three hypotheses, we studied the allele frequencies of the 388 *D. rotundata* sequences, focusing on 144 SNPs that are positioned over the entire genome and are oppositely fixed in the two candidate progenitors (Fig. 2.6B and Fig. 2.7). If Hypothesis 1 or 2 is correct, the allele frequencies in these 144 SNPs should be highly skewed to either of the progenitors. However, the patterns of allele contributions from the two candidate species to *D. rotundata* were almost the same. This result suggests that Hypothesis 3, the admixture origin of Guinea yam, is most likely correct.

We tested the three hypotheses by ∂a∂i (Gutenkunst et al., 2009) with symmetric migration (gene flow) rates, using 15,461 SNPs polarized by *D. alata* (Fig. 2.6A), which showed that Hypothesis 3 had the highest likelihood and the lowest Akaike information criterion (AIC) (Fig. 2.6C and Table 2.3). This result supports the admixture hypothesis, that is, that *D. rotundata* was derived from crosses between *D. abyssinica* and *D. praehensilis*. The parameters estimated by ∂a∂i indicate that the hybridization between *D. abyssinica* and *D. praehensilis* was relatively recent in relation to the divergence between the two wild species. This analysis also indicated that the genomic contributions from *D. abyssinica* and *D. praehensilis* during the hybridization period were approximately 68% and 32%, respectively. Introgression generally results in highly asymmetric genomic contributions from the parental species, whereas hybridization shows symmetric genomic contributions (Folk et al., 2018). The intermediate genomic contributions revealed by this analysis support the hybridization rather than the introgression hypothesis. Our finding is in line with the proposal of hybrid origin

of Guinea yam by D.G. Coursey in 1976 based on morphology (Coursey, 1976b) and supports his speculation that spontaneous hybridization between wild yams could have occurred at the artifactual "dump-heaps" created by people living in the savannah between the forest and the Sahara (Coursey, 1976a).

To evaluate the genetic distances of *D. rotundata* from the two parental species for each chromosome, we calculated F_{ST} values (Wright, 1951) (Fig. 2.6D and Table 2.5). The genetic distances from the two parents varied across the different chromosomes, and the overall genetic distance of *D. rotundata* from *D. abyssinica* was smaller than that from *D. praehensilis* (Table 2.5). Intriguingly, chromosome 11, to which we previously mapped the candidate locus for sex determination (Tamiru et al., 2017), had the shortest genetic distance from *D. abyssinica* and the longest genetic distance from *D. praehensilis* among all chromosomes, indicating that chromosome 11 of *D. rotundata* is highly skewed to *D. abyssinica* (Fig. 2.6D and Table 2.5). Similarly, inter-species divergence is different between the autosomes and sex chromosome of the dioecious plant species *Silene* (Hu & Filatov, 2016).

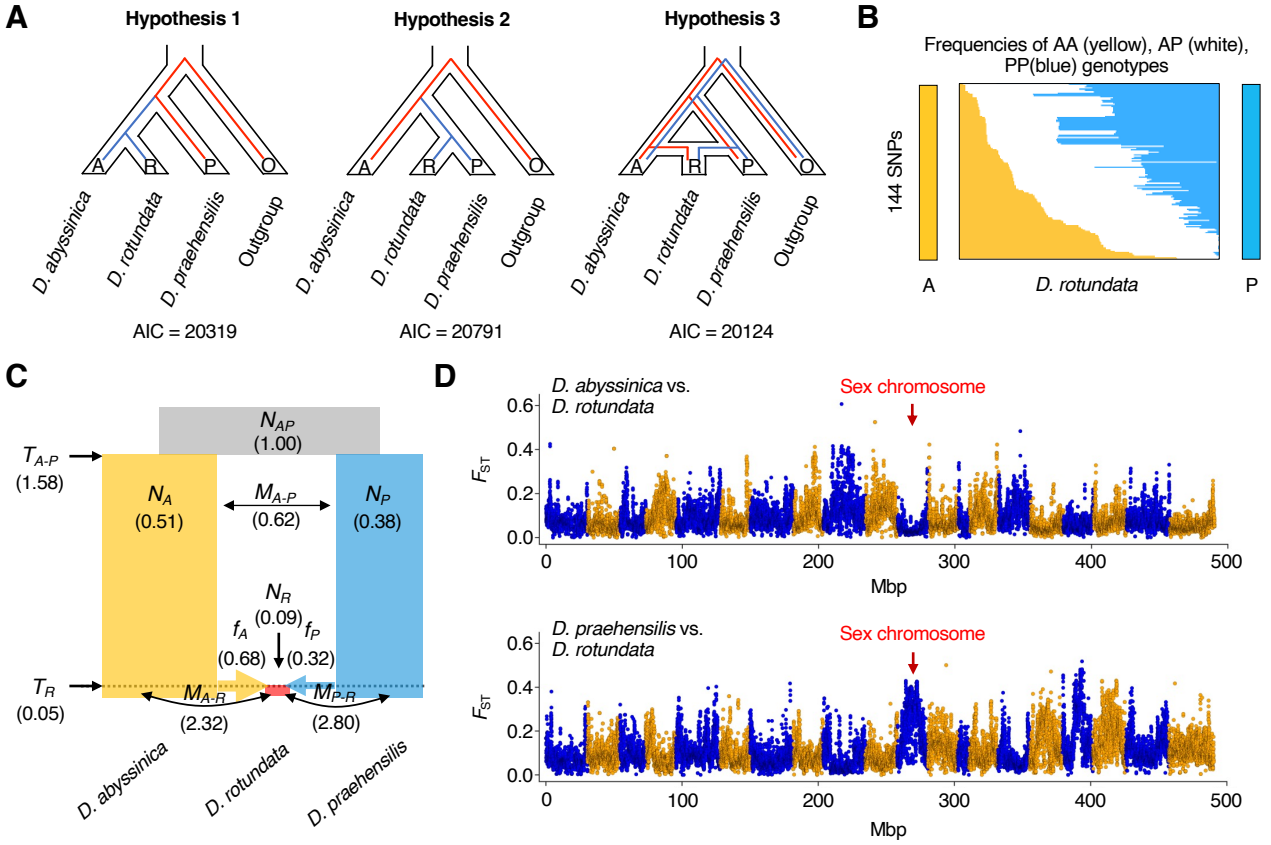


Fig. 2.6. Evidence for the hybrid origin of Guinea yam. (A) Hypotheses for the domestication of Guinea yam (*D. rotundata*). Hypothesis 1 assumes that *D. rotundata* diverged from *D. abyssinica*. Hypothesis 2 assumes that *D. rotundata* diverged from *D. praehensilis*. Hypothesis 3 assumes that *D. rotundata* was derived from a hybrid between *D. abyssinica* and *D. praehensilis*. *D. alata* was used as an outgroup. **(B)** Frequencies of individuals homozygous for *D. abyssinica* allele (A: indicated by yellow color), homozygous for *D. praehensilis* allele (P: indicated by blue color), and heterozygous for A and P (indicated by white color) among the 388 *D. rotundata* sequences as studied for 144 SNPs. **(C)** Evolutionary parameters related to the hybrid origin of Guinea yam as inferred by ∂a∂i (Gutenkunst et al., 2009) using 15,461 SNPs. *N*, *M*, and *T* represent the relative population size from N_{AP}, migration rate, and divergence time, respectively. f_A and f_p indicate the genomic contributions from *D. abyssinica* and *D. praehensilis* when the hybridization occurred, respectively. **(D)** F_{ST} between the wild and cultivated yams. This was conducted with 100-kb window and 20-kb step. Chromosome 11 of *D. rotundata* containing the sex-determining locus shows a lower distance to that of *D. abyssinica* and shows a higher distance to that of *D. praehensilis*.

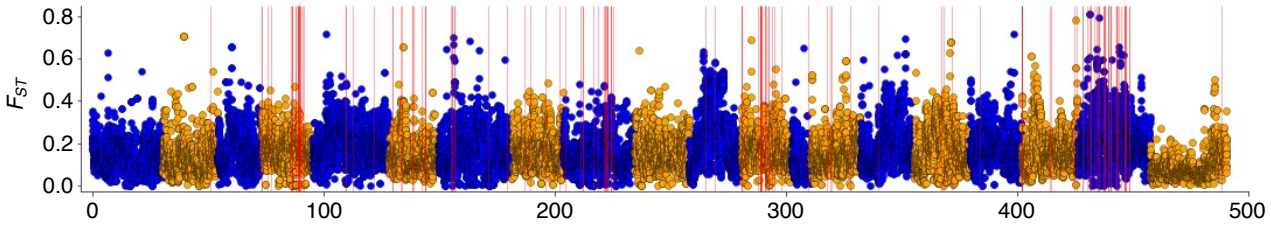


Fig. 2.7. F_{ST} between *D. abyssinica* and *D. praehensilis*. F_{ST} averages were calculated 100-kb window and 20-kb step. The red vertical lines represent the positions of the oppositely fixed SNPs in *D. abyssinica* and *D. praehensilis* as used in Fig. 2.6B.

Table 2.5. F_{ST} in each chromosome. Red and blue indicates the highest and lowest F_{ST} in all chromosomes, respectively. Chromosome 11 of *D. rotundata* containing the sex-determining locus shows a lower distance to that of *D. abyssinica*.

	A vs. P		A vs. R		P vs. R	
Chromosome	F_{ST}	**± std**	F_{ST}	**± std**	F_{ST}	**± std**
All	0.162	0.217	0.082	0.120	0.123	0.157
chrom_01	0.156	0.222	0.079	0.109	0.084	0.112
chrom_02	0.122	0.187	0.055	0.078	0.098	0.121
chrom_03	0.177	0.224	0.075	0.103	0.101	0.115
chrom_04	0.173	0.218	0.111	0.150	0.100	0.130
chrom_05	0.201	0.257	0.098	0.128	0.115	0.133
chrom_06	0.116	0.168	0.065	0.092	0.075	0.102
chrom_07	0.161	0.231	0.093	0.122	0.084	0.114
chrom_08	0.165	0.209	0.120	0.161	0.085	0.109
chrom_09	0.129	0.170	0.129	0.150	**0.062**	0.102
chrom_10	0.152	0.205	0.129	0.169	0.077	0.102
chrom_11	**0.277**	0.273	**0.033**	0.052	**0.247**	0.231
chrom_12	0.160	0.213	0.063	0.096	0.134	0.140
chrom_13	0.111	0.161	0.064	0.100	0.108	0.120
chrom_14	0.141	0.184	0.120	0.163	0.107	0.133
chrom_15	0.204	0.243	**0.133**	0.152	0.073	0.104
chrom_16	0.192	0.248	0.050	0.074	0.174	0.182
chrom_17	0.180	0.201	0.062	0.080	0.217	0.221
chrom_18	0.169	0.210	0.074	0.103	0.188	0.205
chrom_19	0.191	0.240	0.080	0.110	0.133	0.152
chrom_20	**0.070**	0.109	0.057	0.088	0.126	0.143

Evolutionary history of Guinea yam

In angiosperms, plastid genomes are predominantly inherited maternally (McCauley, 1995), making them useful for studying maternal lineages. To infer the maternal history of Guinea yam, we constructed a haplotype network of the whole plastid genome with all samples used in the NJ tree (Fig. 2.2C), as well as the triploid accessions in cluster 1 (Fig. 2.8A and material and method S6). According to this haplotype network, Cameroonian *D. praehensilis* has the largest genetic distance from *D. rotundata*. This result is in line with the phylogenomic trees of African yam (Fig. 2.2C and Fig. 2.2D). Strikingly, the plastid genomes of diploid and triploid *D. rotundata* are uniform and are very similar to that of Nigerian or Beninese *D. abyssinica,* although the latter has another plastid genome lineage distant from that of *D. rotundata*. The plastid genomes of *D. praehensilis* from Nigeria, Benin, and Ghana appear to be derived from Nigerian or Beninese *D. abyssinica*. These results indicate that *D. abyssinica* is an older lineage than *D. praehensilis* and that the places of origin of *D. rotundata* and *D. praehensilis* are probably around Nigeria or Benin. Based on the whole-genome diversity of *D. rotundata*, a recent study (Scarcelli et al., 2019) hypothesized that the origin of *D. rotundata* was around north Benin, as supported by the current results. The plastid genomes of some wild species are identical to those of cultivated Guinea yams. Hybridization between cultivated yams and wild yams may account for this observation (Scarcelli et al., 2017).

The results of nuclear genome admixture (Fig. 2.6) and plastid haplotype network (Fig. 2.8A) analyses indicate that the maternal origin of diploid *D. rotundata* is *D. abyssinica* and its paternal origin is *D. praehensilis* (Fig. 2.8B). Hybridization between *D. abyssinica* and *D. praehensilis* is rare (Scarcelli et al., 2019), but such rare hybrids appear to have been domesticated by humans. The triploid *D. rotundata* shares its plastid haplotype with diploid *D. rotundata*, indicating that diploid *D. rotundata* served as the maternal parent and *D. togoensis* as the paternal parent. *D. cayenensis* is reported to have *D. rotundata* as the maternal parent and *D. burkilliana* as the paternal parent (Girma et al., 2014; Terauchi et al., 1992). All cultivated Guinea yams are hybrids containing *D. abyssinica* plastid genomes.

To explore the changes in population size, we re-inferred the demographic history of African yam by ∂a∂i (Gutenkunst et al., 2009), allowing migration (Fig. 2.8C and material and method S7). We used the same dataset as in Fig. 2.6C. By fixing the parameters predicted in Fig. 2.6C except for population size, we re-estimated each population size at the start and end points after the emergence of these species, assuming an exponential increase/decrease in population size. According to this analysis, since the emergence of the wild progenitors of Guinea yam, the population size of *D. abyssinica* has been decreasing, while that of *D. praehensilis* has been increasing (Fig. 2.8C). This finding suggests that the *D. praehensilis* population was derived from *D. abyssinica*, which is consistent with the results of haplotype network analysis (Fig. 2.8A).

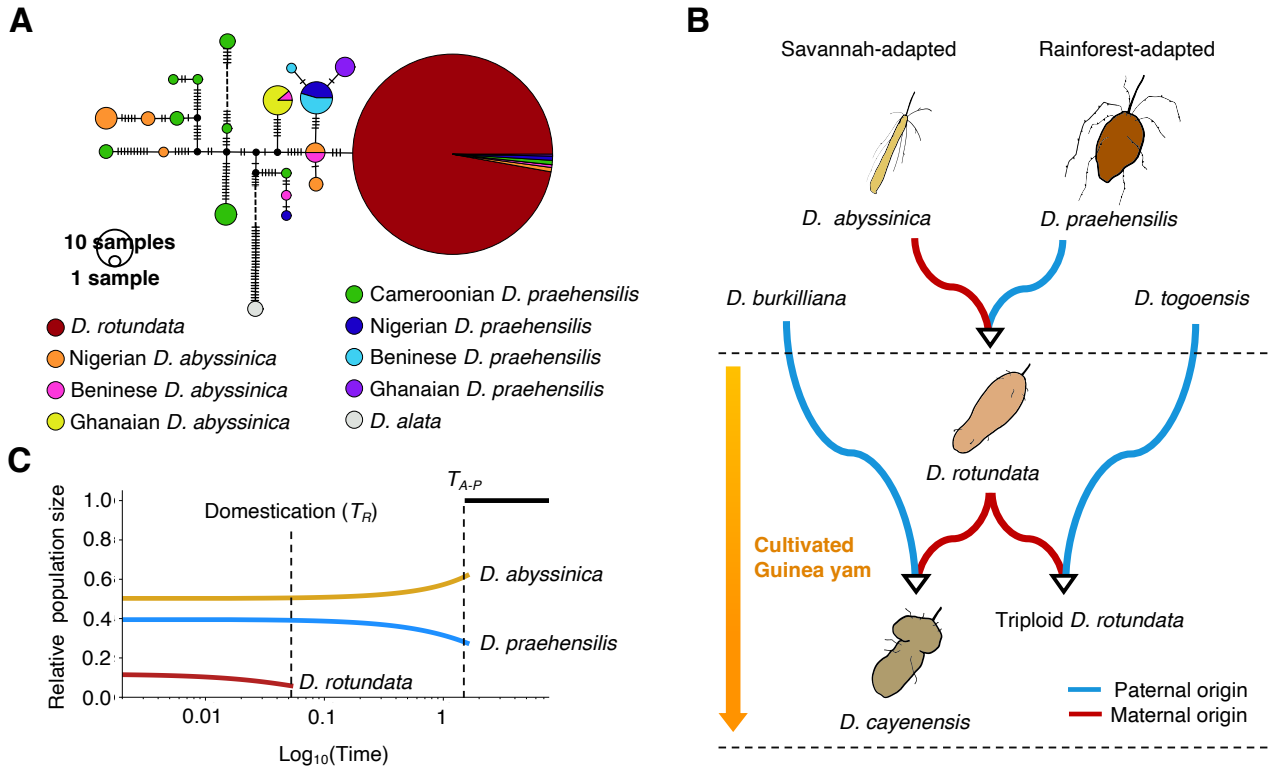


Fig. 2.8. Evolutionary scenario of African yam origins. (A) Haplotype network of the whole plastid genomes of 416 *D. rotundata* (including the triploid accessions), 68 wild relatives, and two *D. alata* accessions used as the outgroup. The number of vertical dashes represents the number of mutations. Western (Nigerian, Beninese, and Ghanaian) *D. praehensilis* and *D. rotundata* seem to have diverged from Nigerian and Beninese *D. abyssinica*. **(B)** Possible scenario of domestication of Guinea yam. The blue line represents paternal origin, and the red line represents maternal origin. **(C)** Changes in population sizes of *D. rotundata* and its wild relatives as inferred by ∂a∂i (Gutenkunst et al., 2009). The parameters except for that of population size were identical to those used in Fig. 2.6C. After the domestication of *D. rotundata*, the population size of *D. rotundata* has increased with migration from the wild progenitors.

Extensive introgression at the *SWEETIE* locus

To explore multiple introgression to *D. rotundata* from the two wild species, we analyzed the f_4 statistic (Reich et al., 2009) using four groups: a) *D. rotundata* cluster 2 and 5; b) *D. rotundata* cluster 4; c) *D. abyssinica*; and d) *D. praehensilis* (material and method S8). The f_4 statistic reveals the representation of two alternative discordant genealogies (Fig. 2.9A). The f_4 value is close to zero if the two groups (group a and b) of *D. rotundata* show a concordant genealogy in relation to *D. abyssinica* and *D. praehensilis*. By contrast, the f_4 value diverges from zero if the two groups of *D. rotundata* exhibit discordant genealogy and a large genetic distance to each other. We obtained the f_4 statistic f_4 (P_{25}, P_4, P_P, P_A) for each SNP and performed sliding window analysis (Fig. 2.9B). The f_4 value was close to zero across the genome, indicating that overall, we cannot decide between topology 1 and 2. However, the genomic regions around the *SWEETIE* gene showed the lowest f_4 (P_{25}, P_4, P_P, P_A) [$Z(f_4)$ = -5.66], with overrepresentation of topology 2 in the *SWEETIE* gene （DRNTG_01731) (Table S3).

To explore the genealogical relationships around the *SWEETIE* gene, we constructed a Neighbor-Net (Huson & Bryant, 2006) around this locus (4.00 to 4.15 Mb on chromosome 17) (Fig. 2.9C). The Neighbor-Net showed that the locus of cluster 4 was close to that of *D. praehensilis*, while the loci of cluster 2 and 5 and some other accessions were close to that of *D. abyssinica*. These results indicate that the *SWEETIE* gene was introgressed from the wild species more than once. The *SWEETIE* gene encodes a membrane protein involved in the general control of sugar flux (Veyres et al., 2008a). The *Arabidopsis thaliana sweetie* mutant shows pronounced changes in the accumulation of sugar, starch, and ethylene along with significant changes in growth and development (Veyres et al., 2008b). We still do not know the effect of this introgression on the phenotype of Guinea yam, but this locus appears to be a target of selection.

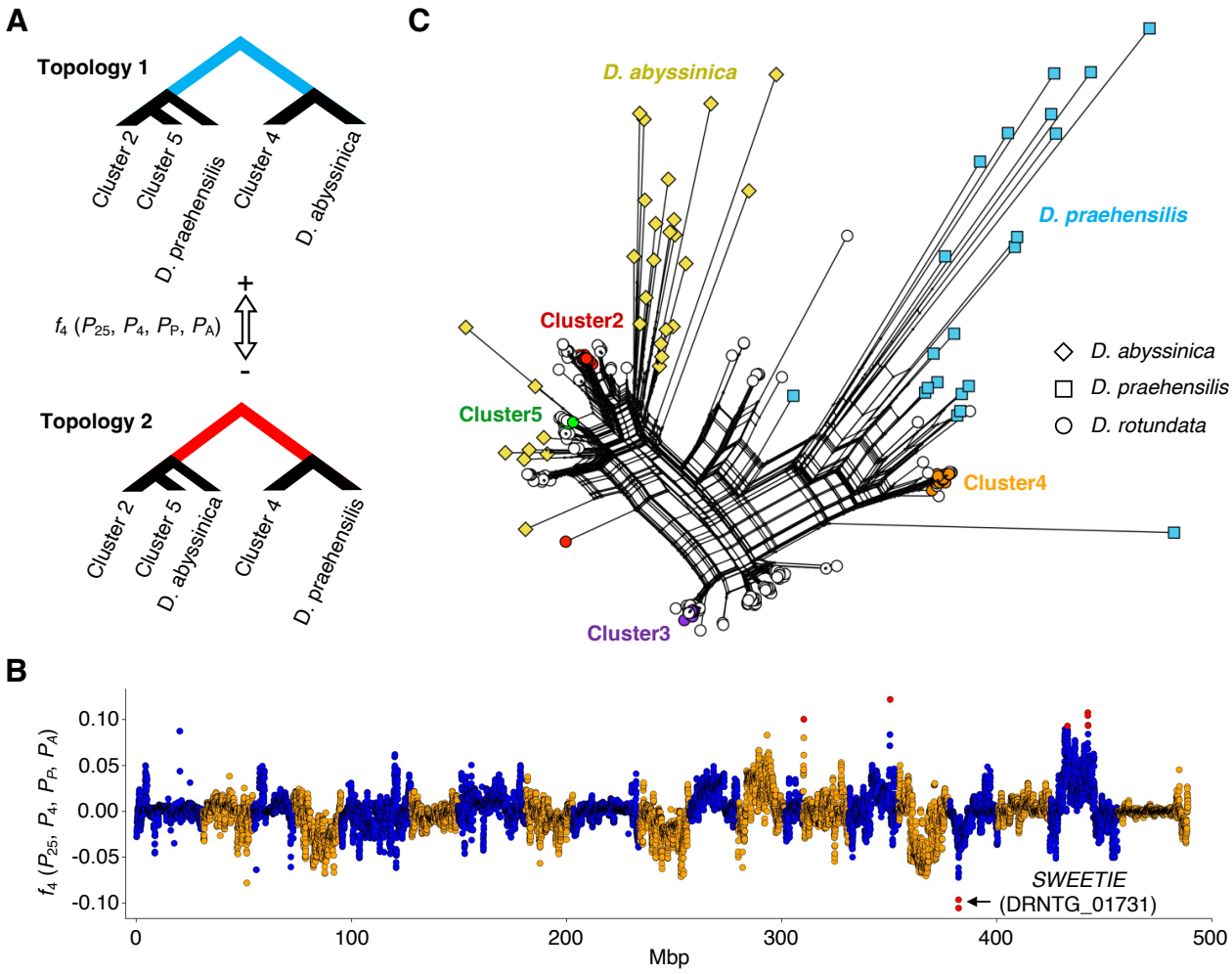


Fig. 2.9. Signature of extensive introgression around the *SWEETIE* gene. (A) Topology of f_4 (P_{25}, P_4, P_P, P_A) in cluster 2, 4, 5 and wild yams. Positive f_4 values represent the long internal branch of the upper tree (Topology 1). Negative f_4 values represent the long internal branch of the bottom tree (Topology 2). **(B)** f_4 values across the genome. This was conducted with 250-kb window and 25-kb step. Red dots indicates outliers of the sliding window which have $|Z(f_4)| > 5$.The locus around the *SWEETIE* gene shows extraordinarily negative f_4 values. **(C)** Neighbor-Net around the *SWEETIE* gene (4 ~ 4.15 Mb on chromosome 17). This was constructed by SplitsTree (Huson & Bryant, 2006) using a total of 458 SNPs.

DISCUSSION

Homoploid hybridization as the trigger of domestication

The importance of hybridization and polyploidization for crop domestication is well documented (Hughes et al., 2007; Salman-Minkov et al., 2016), including in bread wheat (Peng et al., 2011) and banana (Heslop-Harrison & Schwarzacher, 2007). Compared to allopolyploidy, only a limited number of homoploid hybridizations have been reported in plants (Rieseberg, 1991), and homoploid hybridizations have rarely contributed to the origin of crops (Zhang et al., 2019). Homoploid hybridization can increase genetic variation via recombination between distantly related species, and it often allows the hybrid to adapt to unexploited niches (Mallet, 2007). In the case of Guinea yam, the savannah-adapted wild species *D. abyssinica* and the rainforest-adapted wild species *D. praehensilis* are not suitable for agriculture; however, their hybrid, *D. rotundata*, could have been adopted for cultivation by humans. Gene combinations from different wild yams might have contributed to the domestication of Guinea yam. The current study provides an example of the origin of a crop through homoploid hybridization.

Use of wild species to improve Guinea yam

A project for the improvement of Guinea yam by crossbreeding has been initiated (AfricaYam: https://africayam.org). However, the current breeding projects depend predominantly on *D. rotundata* genetic resources. Systematic efforts are needed to introgress beneficial alleles from wild species into crops; these alleles will increase disease resistance and abiotic stress tolerance to improve crop resiliency and productivity (Warschefsky et al., 2014). Our study revealed that the two wild progenitor species (*D. abyssinica* and *D. praehensilis*) of Guinea yam contain much greater genetic diversity than *D. rotundata* (Fig. 2.6C), suggesting that these wild species could be useful sources for alleles of agricultural importance. However, the *D. abyssinica* and *D. praehensilis* accessions in IITA genebank account for only 1.6% of the total *Dioscorea* accessions maintained as of 2018 (Darkwa et al., 2020). Therefore, it will be important to collect and preserve wild *Dioscorea* species as genetic resources for improving Guinea yam. Our findings suggest that new alleles of loci such as the *SWEETIE* gene were introgressed from wild yams into cultivated Guinea yams multiple

times, which likely conferred plants with phenotypes preferred by humans. Many more alleles from wild species remain to be exploited for systematic breeding. Our findings highlight the need to consider how to effectively leverage the gene pools of wild species from different habitats for the rapid breeding of Guinea yam using genomic information.

MATERIALS AND METHODS

S1. Reference assembly

S1.1 Whole-genome sequencing using Oxford Nanopore Technology

S1.2 Quality control

S1.3 *De novo* assembly

S1.4 Polishing and removing duplicated contigs

S1.5 Gene prediction and annotation

S2. Generation of pseudo-chromosomes by anchoring contigs onto a linkage map

S2.1 Preparing the mapping population

S2.2 Whole-genome resequencing

S2.3 Quality control and alignment

S2.4 Identification of parental line-specific heterozygous markers

S2.5 Anchoring and ordering contigs

S3. Genetic diversity analysis

S3.1 Whole-genome resequencing of Guinea yam accessions

S3.2 Quality control, alignment, and SNP calling

S3.3 Unsupervised clustering analysis

S3.4 Polymorphism and ploidy of nuclear genomes

S4. Phylogenomic analysis of African yam

S4.1 Data preparation

S4.2 Neighbor-joining tree

S4.3 Inferring the evolutionary history of wild *Dioscorea* species using ∂a∂i

S4.4 Inferring the evolutionary history of wild *Dioscorea* species using fastsimcoal2

S5. Test of hybrid origin

S5.1 Site frequency spectrum polarized by two candidate progenitors of Guinea yam

S5.2 Inferring the domestication history of Guinea yam using ∂a∂i

S5.3 Comparison of F_{ST} on each chromosome among three African yams

S6. Haplotype network analysis of the whole plastid genome

S7. Inferring the changes in population size

S8. Exploring the possibility of extensive introgression from wild *Dioscorea* specie

S1. Reference assembly

S1.1 Whole-genome sequencing using Oxford Nanopore Technology

To generate version 2 of the *Dioscorea rotundata* reference genome sequence, we sequenced an F1 individual plant named "TDr96_F1" using the PromethION sequencer (Oxford Nanopore Technologies). "TDr96_F1" was the same individual plant used to obtain version 1 of the *D. rotundata* reference genome sequence (Tamiru et al., 2017). "TDr96_F1" DNA was extracted from fresh leaves as described (Tamiru et al., 2017). The DNA was subjected to size selection and purification with a gel extraction kit (Large Fragment DNA Recovery Kit; Zymo Research). The purified DNA was sequenced using PromethION at GeneBay, Yokohama, Japan (http://genebay.co.jp).

S1.2 Quality control

As a first step in our pipeline for genome assembly (Fig. SM1), we removed the lambda phage genome from raw reads with NanoLyse v1.1 (De Coster et al., 2018). We then filtered out reads with an average read quality score of less than 7 and those shorter than 1,000 bases with Nanofilt v2.2 (De Coster et al., 2018). This was followed by trimming of the first 75 bases to remove low-quality bases in all read that were retained. This generated 3,124,439 reads, corresponding to 20.89 Gbp of sequence (Table SM1).

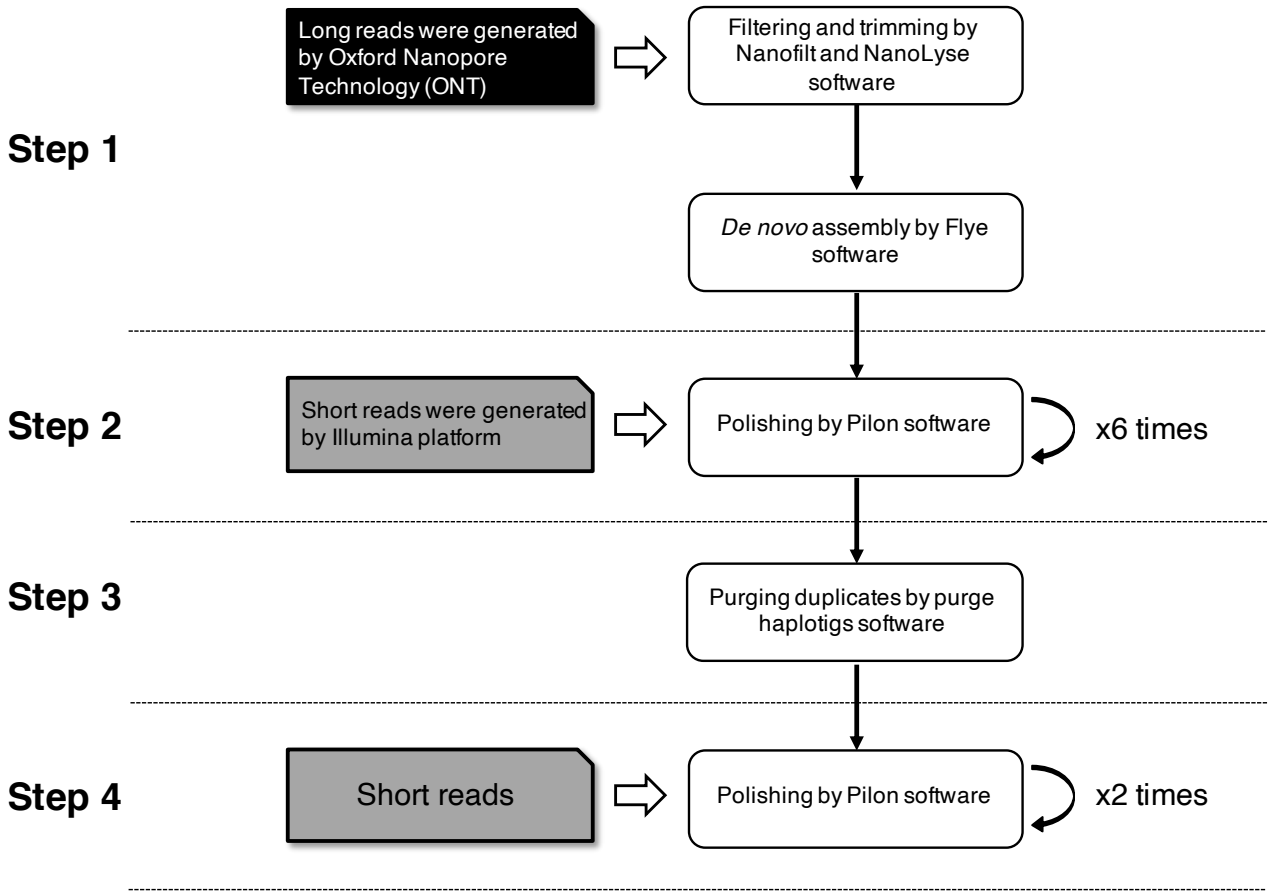


Fig. SM1. Pipeline of genome assembly Ver. 2.

Table SM1. Summary of filtered ONT reads.

Summary	
Number of reads	3,124,439
Total base pairs (Gb)	20.89
Genome coverage	36.6x
Average fragment size (Kb)	6.7
Longest fragment	211,597
Shortest fragment	1,000
Fragment N50 (Kb)	8.0

- Raw reads were registered in DDBJ under accession number DRR196916.

- Genome coverage was estimated based on the expected genome size of *D. rotundata* (570Mb).

S1.3 *De novo* assembly

We assembled filtered long DNA sequence reads with Flye v2.4.2 (Kolmogorov et al., 2019), using 570 Mb as the estimated genome size of *D. rotundata* (Tamiru et al., 2017). This generated 8,721 contigs with N50 of 137,007 base pairs (Step 1 in Table SM2) and a total size of 636.8 Mb, which is larger than the expected *D. rotundata* genome size of 570 Mb. To evaluate the completeness of the gene set in the assembled contigs, we applied BUSCO analysis (Bench-Marking Universal Single Copy) v3.0.2 (Simão et al., 2015). For BUSCO analysis, we set "genome" as the assessment mode and used Embryophyta *odb9* as the database and obtained 40.7% complete BUSCOs (Step 1 in Table SM2).

Table SM2. Summary of the reference assembly.

	Step 1	Step 2	Step 3	**Step 4**
Total number of contigs	8,721	8,721	6,513	**6,513**
Total base-pairs (Mbp)	636.8	628.2	579.7	**579.4**
Average contig size (bp)	73,008	72,029	89,004	**88,961**
Longest contig (bp)	2,301,335	2,267,833	2,267,833	**2,267,326**
Shortest contig (bp)	171	171	171	**171**
N50 (bp)	137,007	134,605	152,963	**152,929**
Complete BUSCOs (%)	40.7	89.9	89.3	**90.1**
Complete and single-copy BUSCOs (%)	39.9	83.9	84.9	**85.7**
Complete and duplicated BUSCOs (%)	0.8	6.0	4.4	**4.4**
Fragmented BUSCOs (%)	8.2	3.2	3.2	**3.1**
Missing BUSCOs (%)	51.1	6.9	7.5	**6.8**

S1.4 Polishing and removing duplicated contigs

To correct the assembled contigs, we repeatedly polished them with Illumina short reads (Table SM3) using Pilon v1.23 (Walker et al., 2014) until there was no further change in the % of complete BUSCOs. We aligned Illumina jump reads as single reads to the assembled contigs using the bwa mem command in BWA v0.7.17 (Li & Durbin, 2009) and sorted the BAM files with SAMtools v1.9 (Li et al., 2009). The BAM files were used to run Pilon with the option "--diploid". We polished the contigs six times. The percentage of complete BUSCOs was 89.9% after the first polishing step (Step 2 in Fig. SM1). To remove duplicated contigs, we used Purge Haplotigs v1.0.2 (Roach et al., 2018), which removes duplicated contigs based on depth and the number of matching bases (Step 3 in Fig. SM1). In Purge Haplotigs, the percent cutoff of aliment coverage was set to 95%. Finally, we polished the contigs again. The percentage of complete BUSCOs was 90.1% after the second polishing process (Step 4 in Fig. SM1). Comparing the features in the old reference genome with the new reference genome, the number of missing bases ("N") was drastically reduced (Table SM4).

Table SM3. Sequence list used for polishing.

Name	Sequence Platform	Total size (Gb)	Genome coverage	Accession No.
Fragment (PE)	Illumina Miseq	16.77	29.4x	DRR027644
MP jump reads (as Single)				
for 2k	Illumina Hiseq 2500	6.43	11.3x	DRR027645
for 3k	Illumina Hiseq 2500	7.56	13.3x	DRR027646
for 4k	Illumina Hiseq 2500	6.18	10.8x	DRR027647
for 5k	Illumina Hiseq 2500	7.20	12.6x	DRR027648
for 6k	Illumina Hiseq 2500	7.27	12.8x	DRR027649
for 8k	Illumina Hiseq 2500	6.79	11.9x	DRR027650

- All values are calculated after quality control.

- Genome coverage was estimated based on the expected genome size of *D. rotundata* (570 Mb).

Table SM4. Comparison of the old (Tamiru et al., 2017) and new reference assemblies.

Feature	Ver. 1	Ver. 2
Number of scaffolds*	4,723	6,513
Total scaffold* size (Mbp)	594.23	579.41
Longest scaffold* (Mbp)	13.61	2.28
N50 (Mbp)	2.12	0.15
Total 'N' bp	90,097,902	953
Complete BUSCOs (%)	90.7	90.1

*In Version 2, contigs were used instead of scaffolds.

S1.5 Gene prediction and annotation

For gene prediction, we used 20 RNA-Seq data sets representing 15 different organs and three different flowering stages in male and female plants (Table SM5). Total RNA was used to construct cDNA libraries using a TruSeq RNA Sample Prep Kit V2 (Illumina) according to the manufacturer's instructions. The extracted RNA was sequenced on the Illumina platforms NextSeq500 and HiSeq4000. In the quality control step, we filtered the reads and discarded reads shorter than 50 bases and those with an average read quality below 20 and trimmed poly(A) sequences with FaQCs v2.08 (Lo & Chain, 2014). Quality trimmed reads were aligned to the newly assembled contigs with HISAT2 v2.1 (Kim et al., 2015) with the options "--no-mixed --no-discordant --dta". Transcript alignments were assembled with StringTie v1.3.6 (Pertea et al., 2015) separately for each BAM file. These GFF files were integrated with TACO v0.7.3 (Niknafs et al., 2017) with the option "--filter-min-length 150", generating 26,609 gene models within the new assembly (Table SM6). Additionally, coding sequences (CDSs) that were predicted using the previous reference genome (Tamiru et al., 2017) were aligned to the newly assembled contigs with Spaln2 v2.3.3 (Iwata & Gotoh, 2012). Consequently, 8,889 CDSs that did not overlap with the new gene models were added to the new gene models (Table SM6). Finally, gene models shorter than 75 bases were removed, and InterProScan v5.36 (Jones et al., 2014) was used to predict ORFs (open reading frames) and strand information for each gene model. We predicted 35,498 genes, including 66,561 transcript variants (Table SM6). For gene annotation, the predicted gene models were searched in the Pfam protein family database using InterProScan (Jones et al., 2014) and with the blastx command in BLAST+ (Camacho et al., 2009) with the option "-evalue 1e-10", using the Viridiplantae database from UniProt as the target database. The resulting gene models and annotations were uploaded to ENSEMBL (http://plants.ensembl.org/Dioscorea_rotundata/Info/Index).

Table SM5. Summary of RNA-seq data used for gene prediction.

Sample name	Fastq size		Sequence platform	Comment	Accession No.
	Original (Gbp)	Filtered (Gbp)			
01_Flowers-rachis-top	4.36	4.28	NextSeq500	Top 2 cm of inflorescence	DRR063119
02_Flowers-rachis-lower	4.96	4.87	NextSeq500	Lower 2 cm of inflorescence	DRR063118
03_Flower-bud	3.52	3.46	NextSeq500	Flower bud	DRR063116
04_Axillary-bud	4.31	4.23	NextSeq500	Axillary bud	DRR063115
05_Leaf	3.26	3.18	NextSeq500	Leaf	DRR045127
06_Petiole	4.47	4.38	NextSeq500	Petiole	DRR063121
07_Pulvinus	4.66	4.58	NextSeq500	Pulvinus	DRR063120
08_Rachis	4.59	4.51	NextSeq500	Rachis	DRR063117
09_Stem	3.45	3.36	NextSeq500	Young_stem	DRR045129
10_Spine	4.51	4.43	NextSeq500	Spine	DRR063123
11_Root	3.62	3.54	NextSeq500	Root	DRR063122
12_Tuber-head	4.72	4.65	NextSeq500	Tuber (head)	DRR063126
13_Tuber-middle	4.06	4.00	NextSeq500	Tuber (middle)	DRR063125
14_Tuber-tail	4.48	4.40	NextSeq500	Tuber (tail)	DRR063124
15_fem_Y917-1	4.12	4.08	HiSeq4000	TDr97_00917 female flower early stage 1	DRR208398
16_fem_Y917-2	4.27	4.23	HiSeq4000	TDr97_00917 female flower early stage 2	DRR208399
17_fem_Y917-3	4.43	4.37	HiSeq4000	TDr97_00917 female flower early stage 3	DRR208400
18_mal_Y777-1	4.48	4.42	HiSeq4000	TDr97_00777 male flower early stage 1	DRR208401
19_mal_Y777-2	3.43	3.40	HiSeq4000	TDr97_00777 male flower early stage 2	DRR208402
20_mal_Y777-3	4.13	4.09	HiSeq4000	TDr97_00777 male flower early stage 3	DRR208403

Table SM6. Summary of gene prediction.

	Contigs (6,513)	**Pseudo Chrom. (01~20)**
No. genes	35,498	30,344
(Total transcript variants)	(66,561)	(57,637)
ORF status		
Complete	22,423	19,502
5′ partial	1,225	1,018
3′ partial	10,385	8,594
Internal	559	465
No ORF	906	765
Prediction software		
TACO (*12*)	26,609	23,335
Spaln2 (*13*)	8,889	7,009

S2. Generation of pseudo-chromosomes by anchoring contigs onto a linkage map

S2.1 Preparing the mapping population

To develop the chromosome-scale TDr96_F1 genome sequence from the assembled contigs, we generated an F1 population containing 156 individuals by crossing two *D. rotundata* breeding lines: TDr04/219 as the female parent (P1) and TDr97/777 as the male parent (P2).

S2.2 Whole-genome resequencing

We extracted each DNA sample from dried *D. rotundata* leaves as described (Tamiru et al., 2017). Libraries for PE short reads were constructed using an Illumina TruSeq DNA LT Sample Prep Kit (Illumina). The PE library was sequenced on the Illumina Hiseq4000 platform. A summary of sequence and alignment information is provided in Table S4.

S2.3 Quality control and alignment

We used FaQCs v2.08 (Lo & Chain, 2014) to remove unpaired reads and adapters. We then filtered out reads shorter than 75 bases or those whose average read quality score was 20 or lower with prinseq-lite v0.20.4 lite (Schmieder & Edwards, 2011). We also trimmed bases whose average read quality score was below 20 from the 5′ end and the 3′ end using the sliding window approach (the window size was five bases, and the step size was one base) in prinseq-lite (Schmieder & Edwards, 2011). Subsequently, we aligned the filtered reads of P1, P2, and F1 progenies to the newly assembled contigs (material and method S1) using the bwa mem command in BWA (Li & Durbin, 2009). After sorting the BAM files, we only retained properly paired and uniquely mapped reads using SAMtools (Li et al., 2009).

S2.4 Identification of parental line-specific heterozygous markers

SNP-type heterozygous markers

SNP-based genotypes for P1, P2, and F1 progenies were obtained as a VCF file. The VCF file was generated as follows: (i) SAMtools v1.5 (Li et al., 2009) mpileup command with the option "-t DP,AD,SP -B -Q 18 -C

50"; (ii) BCFtools v1.5 (Li, 2011) call command with the option "-P 0 -v -m -f GQ,GP"; (iii) BCFtools (Li, 2011) view command with the options "-i 'INFO/MQ≥40, INFO/MQ0F≤0.1, and AVG(GQ)≥10"; and (iv) BCFtools (Li, 2011) norm command with the option "-m+any" (Fig. SM2). We rejected the variants with low read depth (<10) or low genotype quality scores (<10) in the two parents. We regarded variants with low read depth (<8) or low genotype quality scores (<5) in F1 progenies as missing and only retained the variants with low missing rates (<0.3).

Subsequently, only bi-allelic SNPs were selected by the BCFtools (Li, 2011) view command with the option "-m 2 -M 2 -v snps". Referring to the genotypes in the VCF file, heterozygous genotypes called by unbalanced allele frequency (out of 0.4-0.6 in two parents, and out of 0.2-0.8 in F1 progenies) were regarded as missing, and filtering for missing rate (<0.3) was applied again. Finally, a binomial test was performed to reject SNPs affected by segregating distortion in the F1 progenies. This binomial test assumes that the probability of success rate is 0.5 based on the two-side hypothesis, and we regarded variants having *p*-value less than 0.2 as segregation distortion.

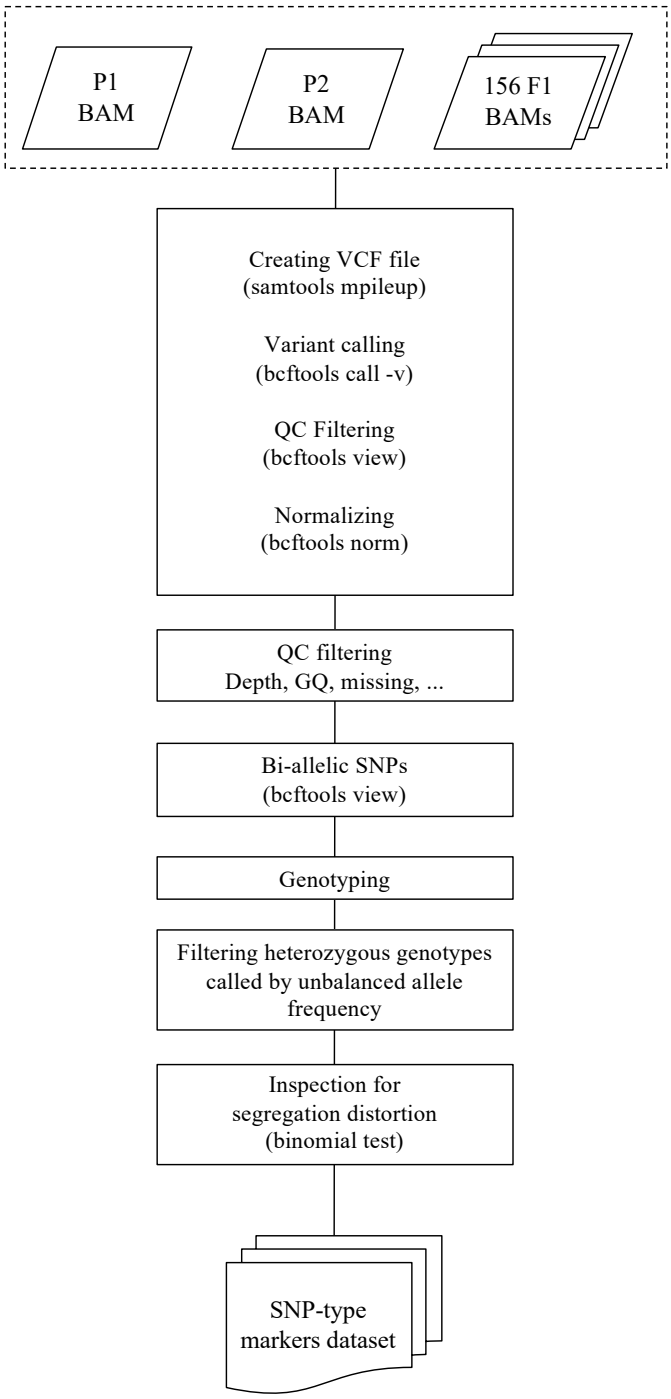


Fig. SM2. Flowchart of SNP-type heterozygous marker selection.

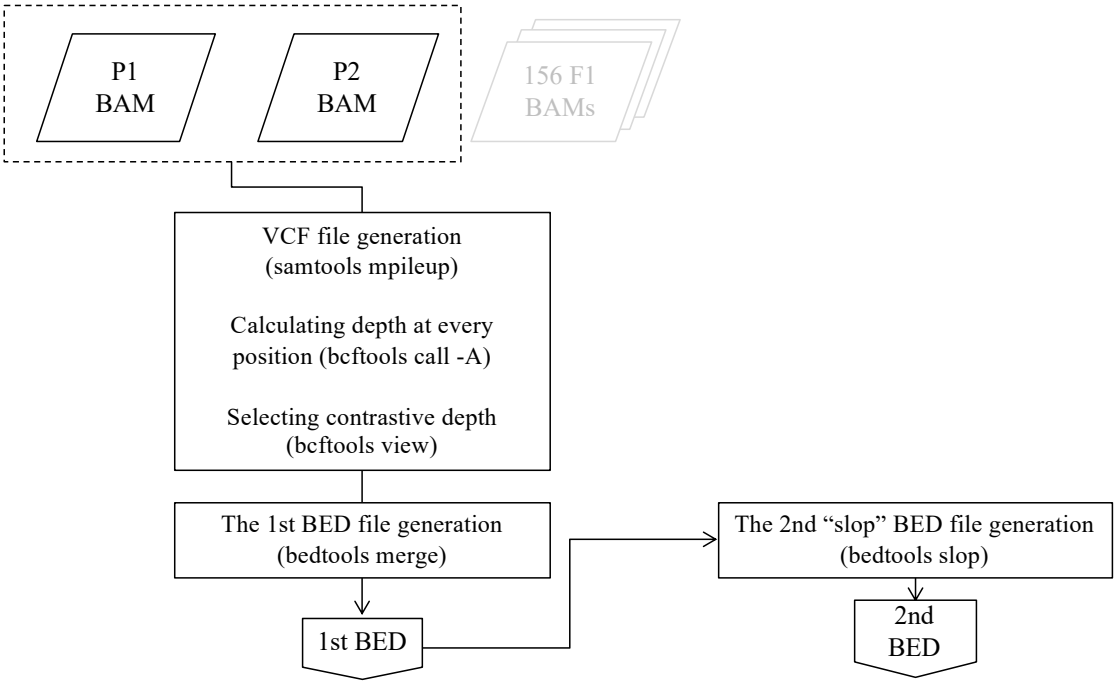
Presence/absence-type heterozygous markers

A VCF file was generated to search for positions with contrasting read depth between the two parental plants P1 and P2 using the following commands: (i) SAMtools (Li et al., 2009) mpileup command with the option "-B -Q 18 -C 50"; (ii) BCFtools (Li, 2011) call command with the option "-A"; and (iii) BCFtools (Li, 2011)

view command with the options "-i 'MAX(FMT/DP)≥8 and MIN(FMT/DP)≤0' -g miss -V indels". This means that one of the parents (P1 or P2) has enough read depth (≥8) and another parent has no reads aligned on that region (A in Fig. SM3). Subsequently, we converted continuous positions in the VCF file to a feature that provides the start and end coordinate information of a region using the BEDTools v.2.26 (Quinlan & Hall, 2010) merge command with the option "-d 10 -c 1 -o count". We only retained sufficiently wide features (≥50 bp) in the BED file (the 1st BED). To reject false positives whereby low-depth regions are erroneously regarded as absent regions, we focused on both the boundary regions around each feature and the features themselves. For boundary regions, the 2nd BED file including expanded (twice-sized) features of each feature given in the 1st BED was generated with the BEDTools (Quinlan & Hall, 2010) slop command with the option "-b 0.5 –pct".

Using the depth value in each feature given in the 1st BED, presence/absence-based genotypes for parental plants P1 and P2 and F1 progenies were determined. To verify the rejection of false-positive features, we also referred to the depth values in the boundary regions around each feature. Verified features were only accepted as presence/absence markers. The depth values in each feature were calculated with the SAMtools (Li et al., 2009) bedcov command with the option "-Q 0". Also, the depth values in the boundary regions were obtained by subtracting the depth values of the 2nd BED from that of the 1st BED (B in Fig. SM3). For P1 and P2, we regarded genotypes having depth ≥ 8 as present genotypes, meaning the heterozygosity of present and absent, while those having depth < 2 were classified as absent genotypes, meaning the homozygosity of absent. For F1 progenies, we classified markers with depth > 0 and = 0 as present and absent markers, respectively. Finally, we applied the same binomial test for SNP-type heterozygous markers as that used for presence/absence-type heterozygous markers.

A


P1
BAM
P2
BAM
156 F1
BAMs
VCF file generation
(samtools mpileup)
Calculating depth at every
position (bcftools call -A)
Selecting contrastive depth
(bcftools view)
The 1st BED file generation
(bedtools merge)
1st BED
The 2nd "slop" BED file generation
(bedtools slop)
2nd
BED

(continued)

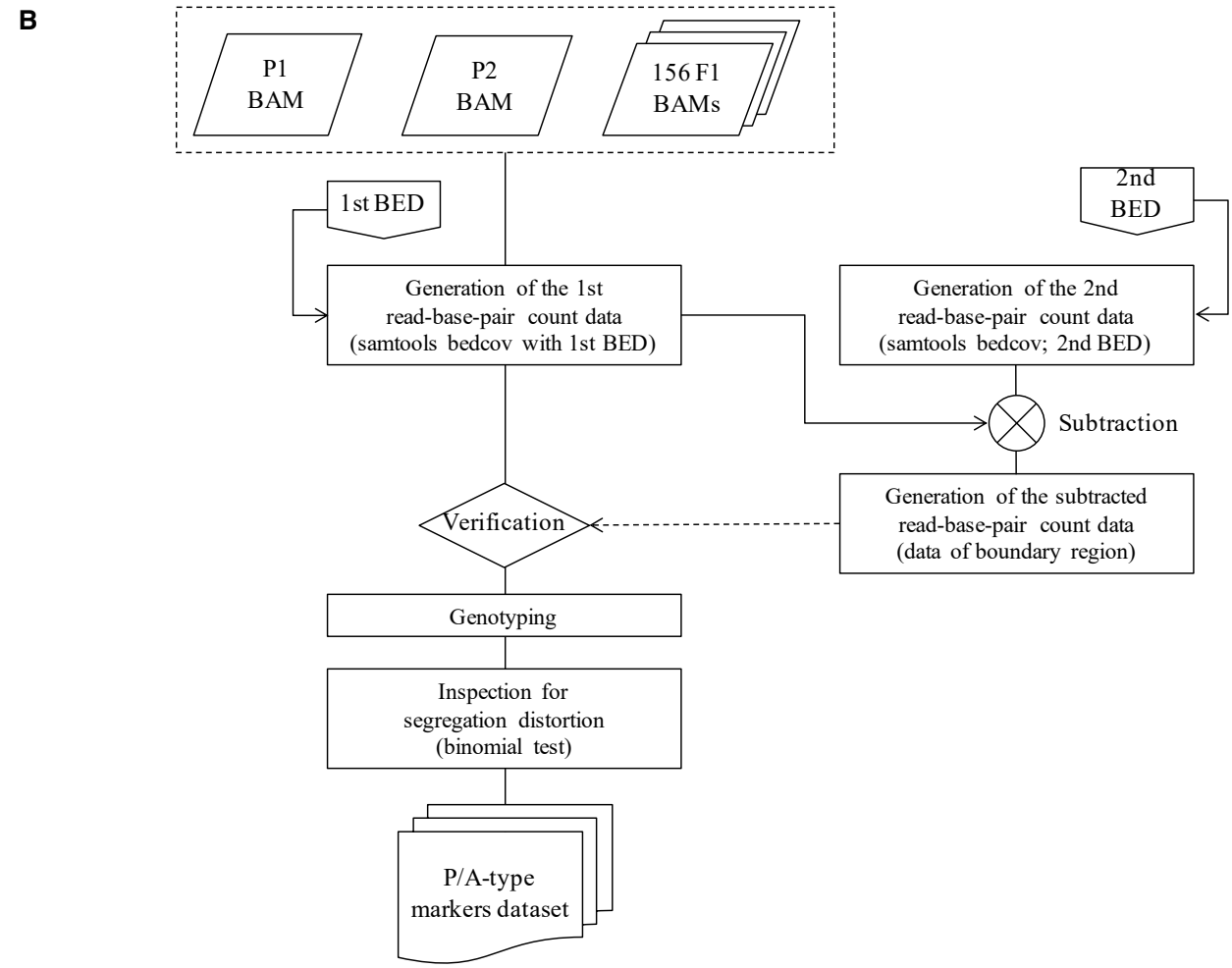


Fig. SM3. Flowchart of presence/absence-type heterozygous marker selection.

Integration of SNP-type and presence/absence-type heterozygous markers

To develop parental line-specific linkage maps, we integrated SNP-type and P/A-type (presence/absence-type) heterozygous markers. Two types of markers were defined: Type-1 markers and Type-2 markers. If an SNP-type marker was heterozygous in P1 but homozygous in P2 or if a P/A-type marker was present in P1 and absent in P2, it was classified as a Type-1 marker (P1-heterozygous marker set). Conversely, if a SNP-type marker was homozygous and heterozygous in P1 and P2, respectively, or if a P/A-type marker was absent in P1 but present in P2, it was classified as a Type-2 marker (P2-heterozygous marker set).

S2.5 Anchoring and ordering contigs

Pruning and flanking markers based on Spearman's correlation coefficients

Distance matrices of Spearman's correlation coefficients (ρ) were calculated for every marker pair in each contig in each marker set (P1-heterozygous marker set and P2-heterozygous marker set). According to the histogram of absolute ρ calculated from each contig, most markers on the same contigs were correlated with each other (Fig. SM4). Therefore, we pruned correlated flanking markers to remove redundant markers (Fig. SM5). Accordingly, we obtained 11,389 markers for linkage mapping (Table SM7).

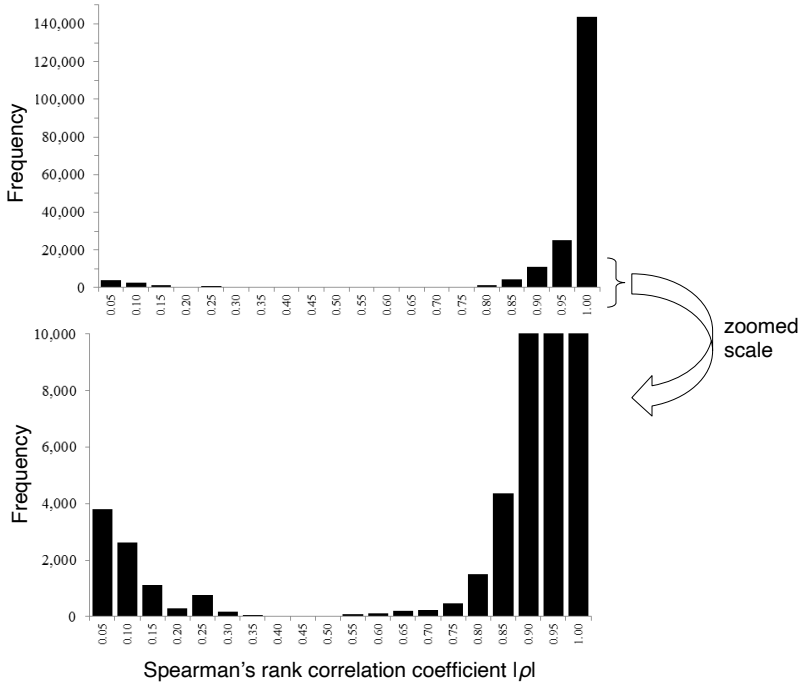


Fig. SM4. Histogram of absolute ρ values calculated from each contig.

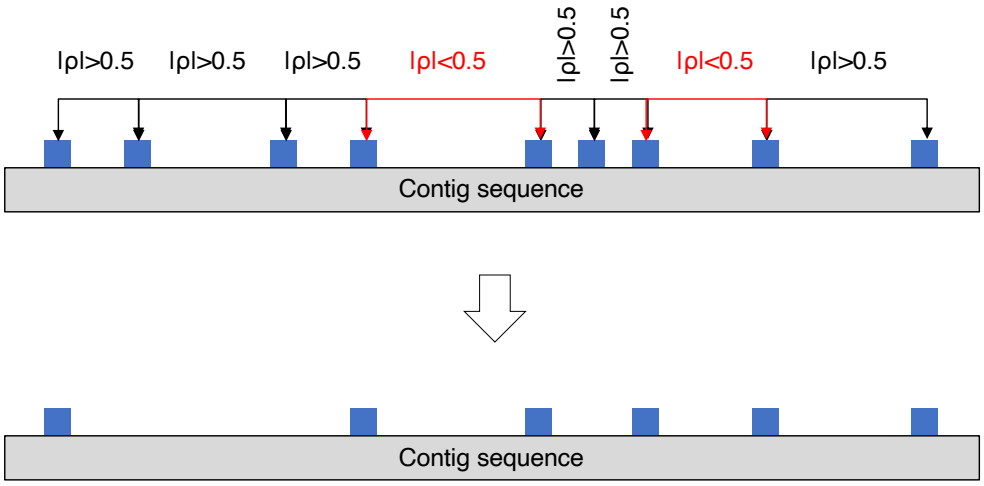


Fig. SM5. The process used to prune correlated flanking markers.

Table SM7. Summary of the anchoring markers.

	Type1	Type2	Type1 + Type2
Total anchoring markers to generate linkage groups	7,020	4,369	11,389
- SNP	4,607	3,435	8,042
- P/A	2,413	934	3,347
Total base pairs of linkage group having markers (Mbp)	434.7	328.4	495.2
Total anchored base pairs estimated from genome size (%)	75.5	56.7	85.5

Linkage mapping

The markers obtained as described in the previous section were converted to genotype-formatted data. Based on this genotype-formatted data, genetic linkage maps were constructed using MSTmap (Wu et al., 2008) with the following parameters: "population_type DH; distance_function kosambi; cut_off_p_value 0.000000000001; no_map_dist 15.0; no_map_size 0; missing_threshold 25.0; estimation_before_clustering no; detect_bad_data no; objective_function ML" for each marker set. After trimming the orphan linkage groups, we solved the complemented-phased duplex linkage groups caused by coupling-type and repulsion-type markers in the pseudo-testcross method. Finally, two parental-specific linkage maps were constructed. These two linkage maps were designated as P1-map (constructed using Type-1 markers) and P2-map

(constructed using Type-2 markers) (Fig SM6 and Fig SM7). The linkage groups were visualized by R/qtl (Broman et al., 2003). The numbering of linkage groups is the same as that used in the previous reference genome (Tamiru et al., 2017).

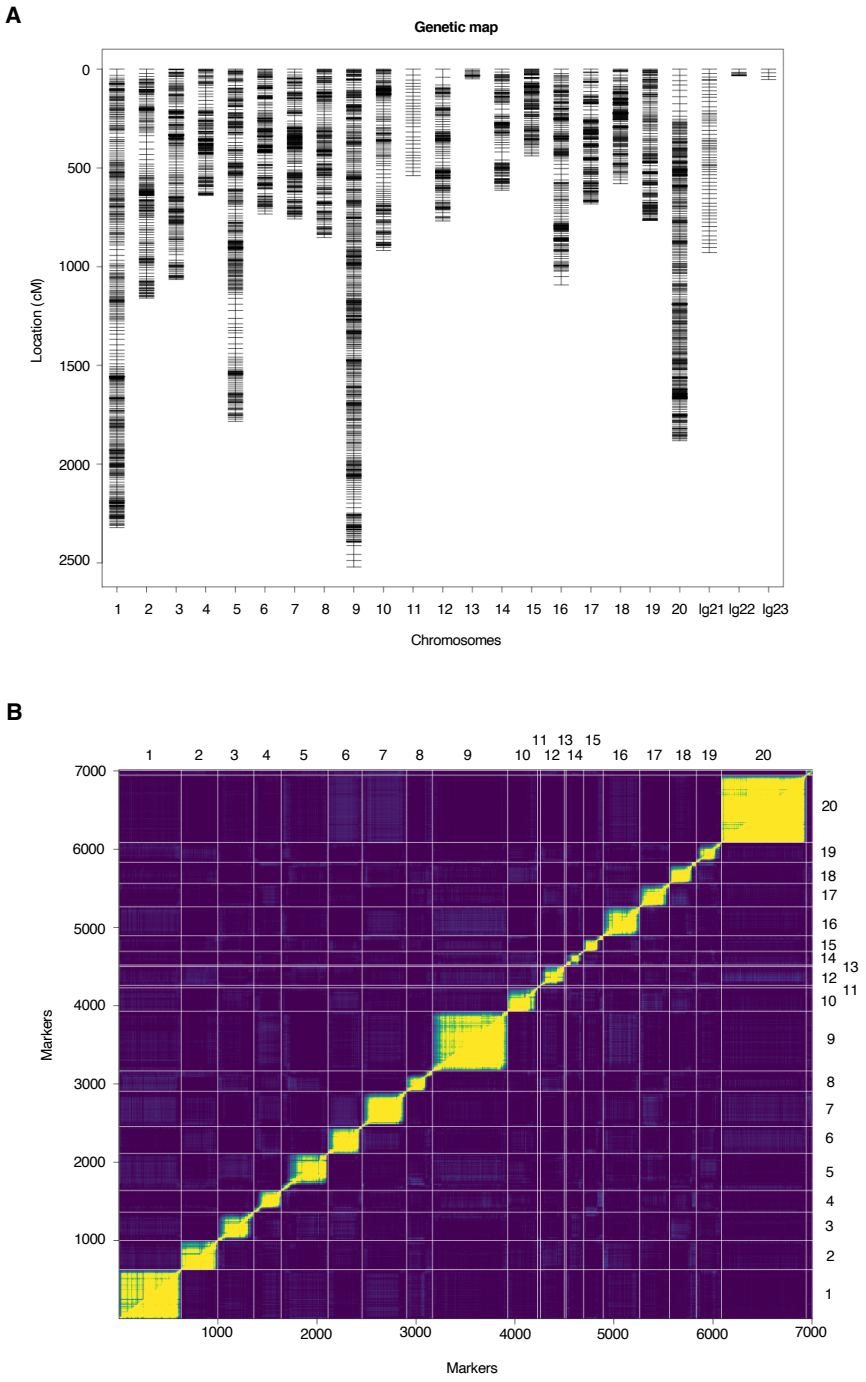


Fig. SM6. P1-map created using P1 heterozygous markers. (A) Contig positions in the P1-map. (B) Estimated recombination fractions (upper-left triangle) against LOD score (lower-right triangle) plotted by R/qtl (Broman et al., 2003).

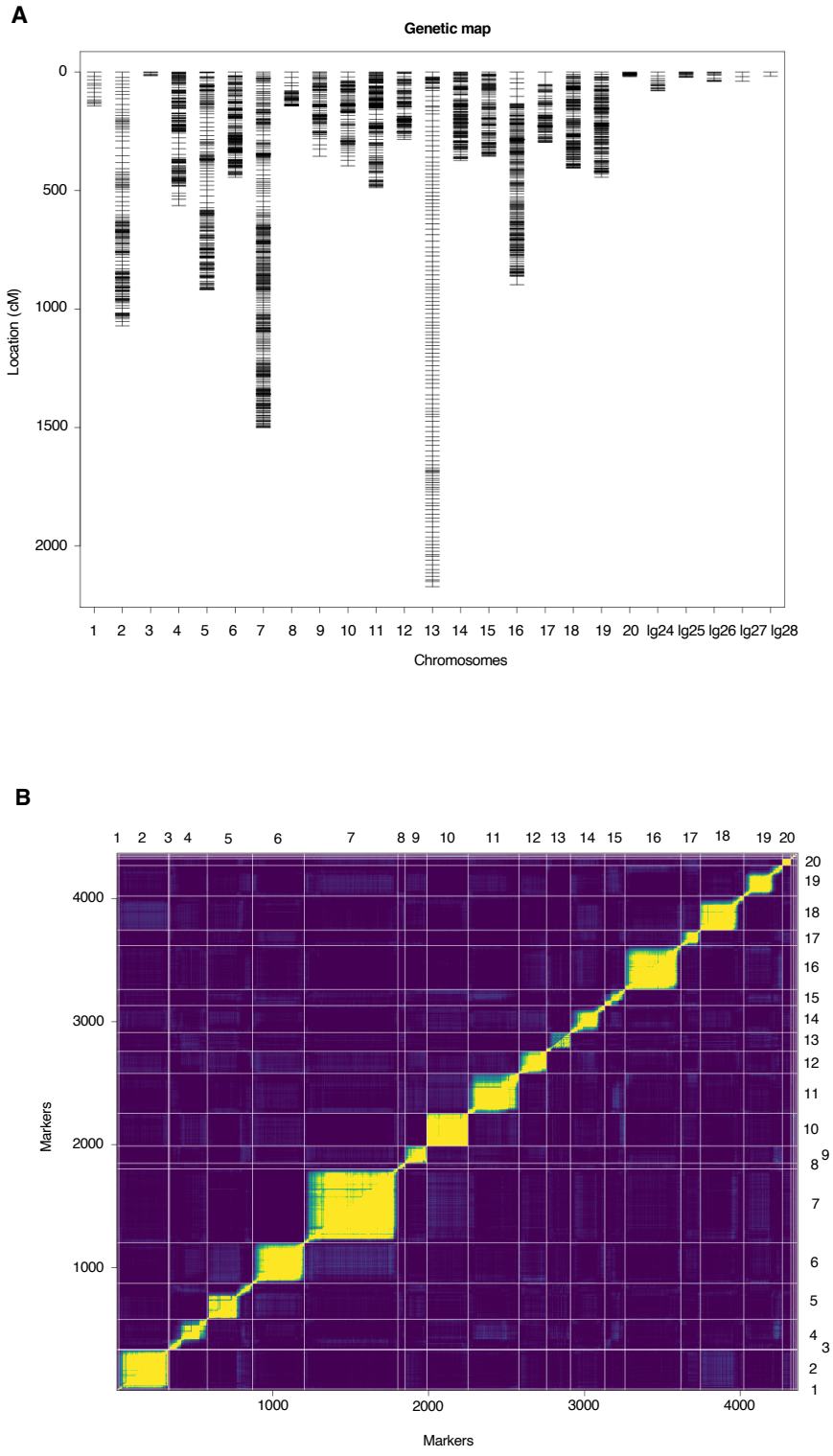


Fig. SM7. P2-map created using P2 heterozygous markers. (A) Contig positions in the P2-map. (B) Estimated recombination fractions (upper-left triangle) against LOD score (lower-right triangle) plotted by R/qtl (Broman et al., 2003).

Integration of two parental-specific linkage maps into the chromosome-scale physical genome sequence

Based on a matrix derived from the contigs shared between the P1- and P2-maps, i.e., linkage groups (Table SM8), the contigs were anchored and linearly ordered as pseudo-chromosomes. During the anchoring and ordering process, we identified contigs whose markers were allocated to different linkage groups. Such contigs were further divided into sub-contigs to ensure that they were not allocated to different pseudo-chromosomes. We divided the contigs at the proper positions as described previously (Tamiru et al., 2017). During this procedure, 34 genes including 61 transcript variants were cut and removed. Finally, a previously described method (Tamiru et al., 2017) was followed to generate the pseudo physical genome sequence composed of 20 pseudo-chromosomes. To compare the newly generated pseudo-chromosomes with the ones we constructed previously (Tamiru et al., 2017), we generated a dot plot with D-Genies (Cabanettes & Klopp, 2018) (Fig. SM8) and counted the anchored base pairs in the new pseudo-chromosomes (Table SM9). The resulting reference genome, including unanchored contigs, was uploaded to ENSEMBL (http://plants.ensembl.org/Dioscorea_rotundata/Info/Index).

Table SM8. A matrix of the number of shared contigs between the P1-map and P2-map. Linkage groups (lg) 21-28 do not have shared contigs.

		P2-map																								
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	lg24	lg25	lg26	lg27	lg28
P1-map	1	5	2	1	2	0	3	2	0	0	3	2	1	0	1	0	5	0	2	0	1	0	0	0	0	0
	2	0	120	0	1	2	2	3	0	1	1	1	0	0	0	0	1	0	1	2	0	0	0	0	0	0
	3	0	2	3	1	0	3	9	0	1	0	0	0	0	0	0	1	0	1	2	0	0	0	0	2	2
	4	0	0	0	84	2	0	1	0	0	0	0	0	0	3	0	1	0	0	0	0	0	0	0	0	0
	5	0	1	0	3	135	2	3	0	1	1	2	2	0	4	1	0	1	1	2	0	0	0	0	0	0
	6	0	0	0	0	3	128	2	0	1	1	2	0	0	1	0	2	0	0	2	0	0	0	0	0	0
	7	0	2	0	1	2	2	199	0	1	1	3	0	0	0	1	1	0	0	3	0	0	0	0	0	0
	8	0	0	0	1	1	4	1	24	0	0	0	0	0	0	1	4	1	2	1	0	9	0	0	0	0
	9	0	1	0	0	2	4	4	0	71	4	1	0	0	2	1	5	1	0	1	0	0	0	0	0	0
	10	0	1	0	0	0	0	1	0	0	93	1	1	0	1	1	0	1	0	0	0	0	6	0	0	0
	11	0	0	0	0	0	0	1	0	0	0	8	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	12	0	0	0	0	2	0	1	0	0	2	2	75	1	0	1	2	0	5	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	1	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	2	1	1	1	0	0	2	0	0	1	66	0	0	1	0	1	1	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	1	0	1	0	0	2	42	2	0	0	1	0	0	0	0	0	0
	16	0	1	0	0	2	0	2	0	2	0	1	1	0	0	0	126	1	1	0	0	0	0	0	0	0
	17	0	0	0	1	2	1	1	0	0	0	1	0	1	1	1	2	60	0	0	0	0	0	0	0	0
	18	0	1	0	0	0	2	1	0	0	1	2	1	0	0	0	0	0	118	0	0	0	0	0	0	0
	19	0	1	0	0	0	1	2	0	0	0	2	0	4	0	0	0	0	1	100	0	0	0	0	0	0
	20	1	8	0	0	5	1	4	0	0	5	6	2	3	2	0	4	1	1	0	39	0	0	3	0	0
	lg21	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	lg22	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	lg23	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

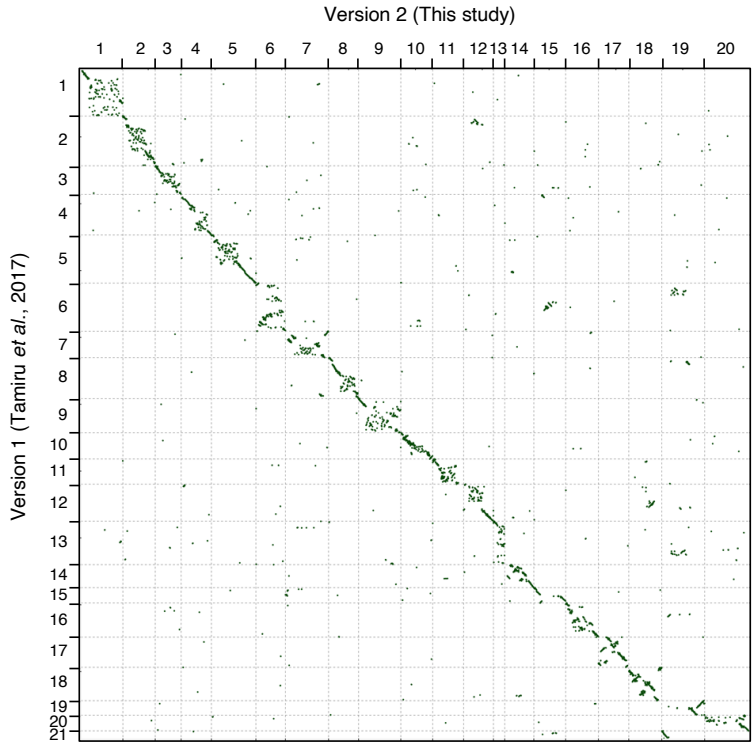


Fig. SM8. Dot plot of the new pseudo-chromosomes (Ver. 2) against the previously generated pseudo-chromosomes (Ver. 1) (Tamiru et al., 2017).

Table SM9. Comparison of the old (Ver. 1) (Tamiru et al., 2017) and new (Ver. 2) pseudo-chromosomes.

Feature	Ver. 1	Ver. 2
Number of Pseudo Chr.	21	20
Total size of Pseudo Chr. (Mbp)	456.67	491.97
Total not 'N' Mbp	406.1	487.31
Total size of Pseudo Chr. / Total scaffold* (%)	76.9	84.9
Complete BUSCOs (%)	82.8%	82.3%

*In version2, contigs were used instead of scaffolds.

S3. Genetic diversity analysis

S3.1 Whole-genome resequencing of Guinea yam accessions

For genetic diversity analysis, we selected 333 accessions of *D. rotundata* maintained at IITA, Nigeria, representing the genetic diversity of Guinea yam landraces and improved lines of West Africa. We extracted DNA from dried leaves of each *D. rotundata* accession as described (Tamiru et al., 2017). Libraries for PE short reads were constructed using an Illumina TruSeq DNA LT Sample Prep Kit (Illumina). The PE library was sequenced on the Illumina Nextseq500 or Hiseq4000 platform. Finally, P1 (TDr04/219) and P2 (TDr97/777) parents used to anchor the contigs and the reference individual "TDr96_F1" were added to the 333 accessions. Therefore, we used a total of 336 accessions for this analysis. A summary of the sequences and alignments is provided in Table S1.

S3.2 Quality control, alignment, and SNP calling

We used FaQCs v2.08 (Lo & Chain, 2014) and prinseq-lite v0.20.4 lite (Schmieder & Edwards, 2011) for quality control. We used the same parameters provided in material and method S2.3, but both paired and unpaired reads were aligned to the new reference genome using the bwa mem command in BWA (Li & Durbin, 2009) with option "-a". After sorting the BAM files, the VCF file was generated using the SAMtools (Li et al., 2009) mpileup command with the option "-t DP,AD,SP -B -Q 18 -C 50", and variants were called by the BCFtools (Li, 2011) call command with the option "-P 0 -v -m -f GQ,GP". Low-quality variants were rejected using the BCFtools (Li, 2011) view command with the options "-i 'INFO/MQ≥40, INFO/MQ0F≤0.1, and AVG(GQ)≥5'". We regarded variants with low read depth (<8) or low genotype quality score (<5) as missing, filtered out SNPs with high missing rates (≥0.3) across all samples, and only retained bi-allelic SNPs on the pseudo-chromosomes.

S3.3 Unsupervised clustering analysis

Through the pipeline described in material and method S3.2, 6,124,093 SNPs were retained in 336 Guinea yam accessions. The VCF file including 336 Guinea yam accessions was converted into a GDS file with the gdsfmt v1.20 R package implemented in the SNPRelate v1.18 (Zheng et al., 2012) R package. We then ran

SNPRelate (Zheng et al., 2012) without filtering for principal component analysis (PCA). Moreover, we used sNMF v1.2 (Frichot et al., 2014) for admixture analysis of the 336 Guinea yam accessions. To choose the best *K* value, we launched sNMF (Frichot et al., 2014) for each *K* value from 2 to 20 (Fig. SM9). We could not find the best *K* value based on the cross-entropy criterion, so we defined five clusters for convenience.

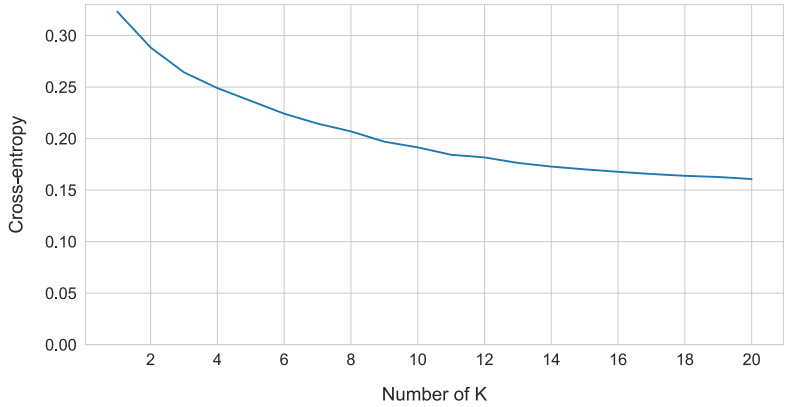


Fig. SM9. Cross-entropy values from *K*=1 to *K*=20 for admixture analysis.

S3.4 Polymorphism and ploidy of nuclear genomes

Heterozygosity level and unique alleles

First, we calculated the heterozygosity level in each accession (Fig. 2.2). We defined the heterozygosity level as follows:

$$(Heterozygosity\ level) = \frac{S}{L}$$

where S is the number of heterozygous SNPs and L is the total number of mapped sites in an accession. The heterozygosity levels of each cluster were statistically compared by two-tailed Student t test (Table 2.1). Second, we counted the unique alleles in each cluster (Fig. 2.3). An allele was considered unique if it only existed in a cluster even when the allele was a singleton in all accessions.

Flow cytometry

Ploidy level was estimated by flow cytometry using a Partec Ploidy Analyzer (Sysmex Partec, Gorlitz, Germany). Fully developed fresh young leaves were sampled and chopped with a razor blade (ca. 5 x 5 mm) in 0.4 mL nuclear extraction buffer (solution A of a High-resolution kit; Sysmex Partec, Gorlitz, Germany). The suspension was filtered through a nylon filter (50-μm mesh), and the extracted nuclei were stained with 4′,6-diamino-2-phenylindole solution. After 5 min of incubation at room temperature, the sample was examined in a ploidy analyzer at a rate of 5–20 nuclei/s. The DNA index (DI) of each accession was calculated based on the relative amount of DNA in nuclei at the G1 stage compared to the internal standard. Rice (*Oryza sativa* L.) was used as an internal standard for calibration of the measurements. Flow cytometry was repeated two or three times with different leaf samples to confirm the DI of each accession. The ploidy levels of each accession were determined by comparing their DI with that of the diploid accession “TDr1673”, for which the chromosome number was confirmed microscopically to be 2n = 40. (Table S1)

Summary statistics of population genetics

After removing the triploid accessions of cluster 1, we imputed missing genotypes using BEAGLE v4.1 (Browning & Browning, 2007) with default options. We then calculated the summary statistics of population genetics (Table 2.2). First, we counted segregating sites and singletons in 308 Guinea yam accessions. We also estimated Watterson's θ ($\hat{\theta}_W$) (Watterson, 1975), pairwise nucleotide diversity ($\hat{\theta}_\pi$) (Nei & Tajima, 1981), and Tajima's *D* (Tajima, 1989) in the same dataset. We defined $\hat{\theta}_W$ as follows:

$$\hat{\theta}_W = \frac{S}{a * \bar{L}}$$

where *a* is equal to:

$$a = \sum_{i=1}^{n-1} \frac{1}{i}$$

and $\bar{L}$ is the number of average mapped sites in a population and *n* is the number of sequences.

We also defined $\hat{\theta}_\pi$ as:

$$\hat{\theta}_\pi = \frac{1}{\bar{L}} \frac{n}{n-1} \frac{\sum_{i<j} k_{ij}}{n(n-1)/2}$$

where $\bar{L}$ is the number of average mapped sites in a population, *n* is the number of sequences, and k_{ij} is the number of nucleotide differences between the *i*th and *j*th sequences.

We also calculated LD decay of 308 Guinea yam accessions (Fig. 2.5). The SNPs whose minor allele frequencies less than 0.05 were removed from the above SNP set used to calculate θ. LD decay was calculated with 200-kb window and 100-kb step. Ten SNPs were randomly sampled within a window, and all possible combinations of r^2 were calculated using the sampled SNPs within a window.

S4. Phylogenomic analysis of African yam

S4.1 Data preparation

For phylogenomic analysis of African yam, we used 308 Guinea yam accessions sequenced in the present study (excluding cluster 1 triploid accessions), as well as 80 *D. rotundata*, 29 *D. abyssinica*, 21 Western *D. praehensilis*, and 18 Cameroonian *D. praehensilis* accessions that were sequenced in a previous study (Scarcelli et al., 2019) using two accessions of the Asian species *D. alata* as an outgroup (Table SM9). Of the samples sequenced in the previous study (Scarcelli et al., 2019) , we only used sequences whose species labels matched a species predicted by admixture analysis in the previous study (Scarcelli et al., 2019) . Also, we removed the sequences that were labeled as hybrids in the previous study (Scarcelli et al., 2019) . Two sequences of *D. alata* downloaded from NCBI were used as the outgroup (Table SM9). Subsequently, read quality control, alignment, and SNP calling of these 458 sequences were conducted using the pipeline described in material and method S3.2. Except for the Neighbor-joining (NJ) tree (Saitou & Nei, 1987) (material and method S4.2), we only used SNPs with a missing rate < 0.3 in each targeted species. When the markers were polarized by comparison with the *D. alata* outgroup, the SNPs at positions where the alleles of *D. alata* were not completely fixed or where either of the *D. alata* sequences was missing were filtered out.

S4.2 Neighbor-joining tree

Before constructing the NJ tree (Saitou & Nei, 1987), we only retained SNPs at positions with no missing data across all five species (*D. rotundata*, *D. abyssinica*, Western *D. praehensilis*, Cameroonian *D. praehensilis*, and *D. alata*). When we converted the VCF file including the remaining SNPs to a multi-FASTA file, heterozygous SNPs were converted to IUPAC code to characterize them as ambiguous markers. To construct the NJ tree (Saitou & Nei, 1987), we ran MEGA X v10.1.8 (Kumar et al., 2018) using the 463,293 remaining SNPs. In MEGA X (Kumar et al., 2018), the bootstrap value was set to 100 and the other parameters were set as default. Finally, the NJ tree was drawn with GGTREE v2.0.4 (Yu et al., 2017).

S4.3 Inferring the evolutionary history of wild *Dioscorea* species using ∂a∂i

To elucidate the evolutionary relationships of the three wild *Dioscorea* species, *D. abyssinica* (indicated as A), Western *D. praehensilis* (P), and Cameroonian *D. praehensilis* (C), which are closely related to *D. rotundata*, we performed ∂a∂i analysis (Gutenkunst et al., 2009). This technique allows evolutionary parameters to be estimated based on an unfolded site frequency spectrum. The joint unfolded site frequency spectrum was calculated based on the 17,532 polarized SNPs and was projected down to 25 chromosomes in each species.

First, three phylogenetic models, {{A, P}, C}, {{P, C}, A}, and {{C, A}, P}, were tested without considering migration among the species. The parameter bounds of each population size ranged from 10^{-3} to 100, and those of each divergence time ranged from 0 to 3, as suggested in the ∂a∂i manual (https://dadi.readthedocs.io/en/latest/). The grid size was set to (40, 50, 60). The maximum iteration for an inference was set to 20. Randomly perturbing the initial values using the 'perturb_params' function in ∂a∂i (Gutenkunst et al., 2009), the parameters were inferred 100 times. Under these conditions, the {{A, P}, C} model had the highest likelihood out of the three models (Table 2.3).

Based on the assumption that {{A, P}, C} represents the true evolutionary relationship among the three wild *Dioscorea* species, the evolutionary parameters were re-estimated by ∂a∂i (Gutenkunst et al., 2009), allowing symmetric migration among the species. The parameter bounds of each symmetric migration rate ranged from 0 to 20, as also suggested in the ∂a∂i manual. The parameters were inferred 100 times by ∂a∂i (Gutenkunst et al., 2009) with different initial parameters, and the best parameter set was selected based on Akaike information criterion.

S4.4 Inferring the evolutionary history of wild *Dioscorea* species using fastsimcoal2

To complement our results and to exactly replicate the conditions used in the previous report (Scarcelli et al., 2019), fastsimcoal2 (Excoffier et al., 2013), which was used in the previous study (Scarcelli et al., 2019), was also used to test these three models ({{A, P}, C}, {{P, C}, A}, and {{C, A}, P}). Until the SNP calling step, we basically followed our own pipeline in material and method S3.2 based on the reference genome version 1 including the unanchored contigs (Tamiru et al., 2017) to be consistent with the previous study (Scarcelli et al., 2019). The misclassified samples excluding hybrids were genetically re-classified by admixture analysis following the methods used in the previous study (Scarcelli et al., 2019). The threshold of missing rate across all samples was set to 0.25, as proposed in the previous study (Scarcelli et al., 2019). We obtained 87,671

SNPs using our pipeline, fewer than the number of SNPs analyzed in the previous coalescent simulation (Scarcelli et al., 2019). Therefore, we skipped the down-sampling of the SNPs to 100,000, unlike in the previous study (Scarcelli et al., 2019). For the other steps and the parameter bounds for the coalescent simulation by fastsimcoal2 (Excoffier et al., 2013), we followed the method used in the previous study exactly (Scarcelli et al., 2019) using the same version of fastsimcoal2 (Excoffier et al., 2013).

S5. Test of hybrid origin

S5.1 Site frequency spectrum polarized by two candidate progenitors of Guinea yam

We focused on the allele frequencies of 388 *D. rotundata* sequences, including 80 from the previous study (Scarcelli et al., 2019), at the SNPs positioned over the entire genome that are oppositely fixed in the two candidate progenitors. The SNP set was generated as described in material and method S4.1. Based on this SNP set, 144 SNPs were oppositely fixed in the two candidate progenitors across all pseudo-chromosomes; the allele frequencies of these 144 SNPs were calculated and plotted.

S5.2 Inferring the domestication history of Guinea yam using ∂a∂i

To infer the domestication history of Guinea yam, we used ∂a∂i (Gutenkunst et al., 2009). Using the 15,461 polarized SNPs generated by following the method in material and method S4.1, three phylogenetic models, {{A, R}, P}, {{P, R}, A}, and {{A, R}, {P, R}} (hypothesis 1, 2, and 3 in Fig. 2A, respectively) were tested, considering symmetric migration among the species. The parameter bound for the admixed proportion from *D. abyssinica* ranged from 0 to 1. The other parameter bounds were the same as in material and method S4.3. The maximum iteration for an inference was set to 20. The parameters were inferred 100 times by ∂a∂i (Gutenkunst et al., 2009).

S5.3 Comparison of F_{ST} on each chromosome among three African yams

F_{ST} (Wright, 1951) among the three species (*D. abyssinica*, [Western] *D. praehensilis*, and *D. rotundata*) was calculated in each chromosome. We estimated F_{ST} using the formula:

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

where H_T and H_S are the expected heterozygosity in the total population and sub-divided population, respectively, which are equal to:

$$H_T = 2\frac{f_{A1} + f_{A2}}{2}(1 - \frac{f_{A1} + f_{A2}}{2})$$

$$H_S = \frac{2f_{A1}(1 - f_{A1}) + 2f_{A2}(1 - f_{A2})}{2} = f_{A1}(1 - f_{A1}) + f_{A2}(1 - f_{A2})$$

where f_{A1} and f_{A2} are the allele frequencies in each population (Wright, 1951). Finally, the calculated F_{ST} were averaged in each chromosome.

S6. Haplotype network analysis of the whole plastid genome

The sample set used to construct the haplotype network of the whole plastid genome was the same as that used to construct the NJ tree (material and method S4.2). We aligned the 458 whole-genome sequences, together with the whole plastid genome of *D. rotundata* (Tamiru et al., 2017), to the newly improved reference genome of *D. rotundata*. We followed the pipeline described in material and method S3.2 for quality control and alignment. Because the plastid genome is haploid, the "--ploidy" option was set to 1 in the BCFtools call command (Li, 2011) when SNPs were called. Singleton SNPs were removed as unreliable markers. SNPs with more than one low-quality genotype (GQ<127) across the samples were also removed as unreliable markers. We did not allow any missing data. Finally, a haplotype network was constructed using the retained 250 SNPs by the median joining network algorithm (Bandelt et al., 1999) implemented in PopART (Leigh & Bryant, 2015).

S7. Inferring the changes in population size

To explore the changes in population sizes, the demographic history of African yams was re-inferred by ∂a∂i (Gutenkunst et al., 2009) allowing migration. By fixing the parameters predicted in material and method S5.2 except for population sizes, we re-estimated each population size at the start and end points after the emergence of these species, assuming an exponential increase/decrease in population size. The parameter bounds of population sizes ranged from 10^{-3} to 100, and the maximum iteration for an inference was set to 20. The parameters were inferred by ∂a∂i 100 times (Gutenkunst et al., 2009).

S8. Exploring the possibility of extensive introgression from *Dioscorea* species

To explore the possibility of multiple introgressions from both parental wild yams, the f_4 statistic (Peter, 2016; Reich et al., 2009) was applied to the four clusters of *D. rotundata* excluding the cluster 1 triploid accessions. Here, calculation of the f_4 statistic requires four populations: P_{R1} is the first cluster of *D. rotundata*, P_{R2} is the second cluster of *D. rotundata*, P_P is a population of (Western) *D. praehensilis*, and P_A is a population of *D. abyssinica*. We estimated $\hat{f}_4(P_{R1}, P_{R2}, P_P, P_A)$ with the following formula using sliding window analysis with a window size of 250 kb and a step size of 25 kb:

$$\hat{f}_4(P_{R1}, P_{R2}, P_P, P_A) = (\hat{p}_{R1} - \hat{p}_{R2})(\hat{p}_P - \hat{p}_A)$$

where $\hat{p}_j$ is the observed allele frequency in a window in population P_j.

In most windows, $\hat{f}_4$ is close to zero, which means that the window has a concordant genealogy because the two clusters of *D. rotundata* have a small genetic distance (B in Fig. SM10). However, if these two clusters of *D. rotundata* have a large genetic distance and if one or both populations have a small genetic distance from a wild *Dioscorea* species, then $\hat{f}_4$ skews from 0. Therefore, a locus having a skewed $\hat{f}_4$ has a discordant genealogy (C or D in Fig. SM10). For P_P (the population of *D. praehensilis*) and P_A (the population of *D. abyssinica*), the samples sequenced in the previous study (Scarcelli et al., 2019) were used (Table SM9), and the dataset was prepared as described in material and method S4.1. As the first screening, all possible combinations of the clusters of *D. rotundata*, excluding accessions in cluster 1, were used for P_{R1} and P_{R2} (Fig. SM11). In this analysis, we identified an extensive introgression around the *SWEETIE* gene (4.00 to 4.15 Mb on chromosome 17). Because clusters 2 and 5 have the same genealogy pattern around the *SWEETIE* gene, we merged them into one population (P_{25}) and used this as P_{R1}. Because cluster 4 has the opposite genealogy pattern to P_{25} around the *SWEETIE* gene, we used P_4 as P_{R2}. As a result, $\hat{f}_4(P_{25}, P_4, P_P, P_A)$ was calculated for the second screening (Fig. 4). If a locus had $|Z(f_4)|>5$, we regarded it as an outlier (red dots in Fig. 4B). To reveal the relationships of the *D. rotundata* accessions around the *SWEETIE* gene, a Neighbor-Net was constructed by SplitsTree v5.1.4 (Huson & Bryant, 2006) using 308 *D. rotundata* accessions excluding the accessions in cluster 1, 29 *D. abyssinica* accessions, and 21 *D. praehensilis* accessions. A total 458 SNPs from

the 4.00–4.15 Mb region on chromosome 17 were converted to multi-FASTA format. At that time, heterozygous genotypes were converted to IUPAC codes.

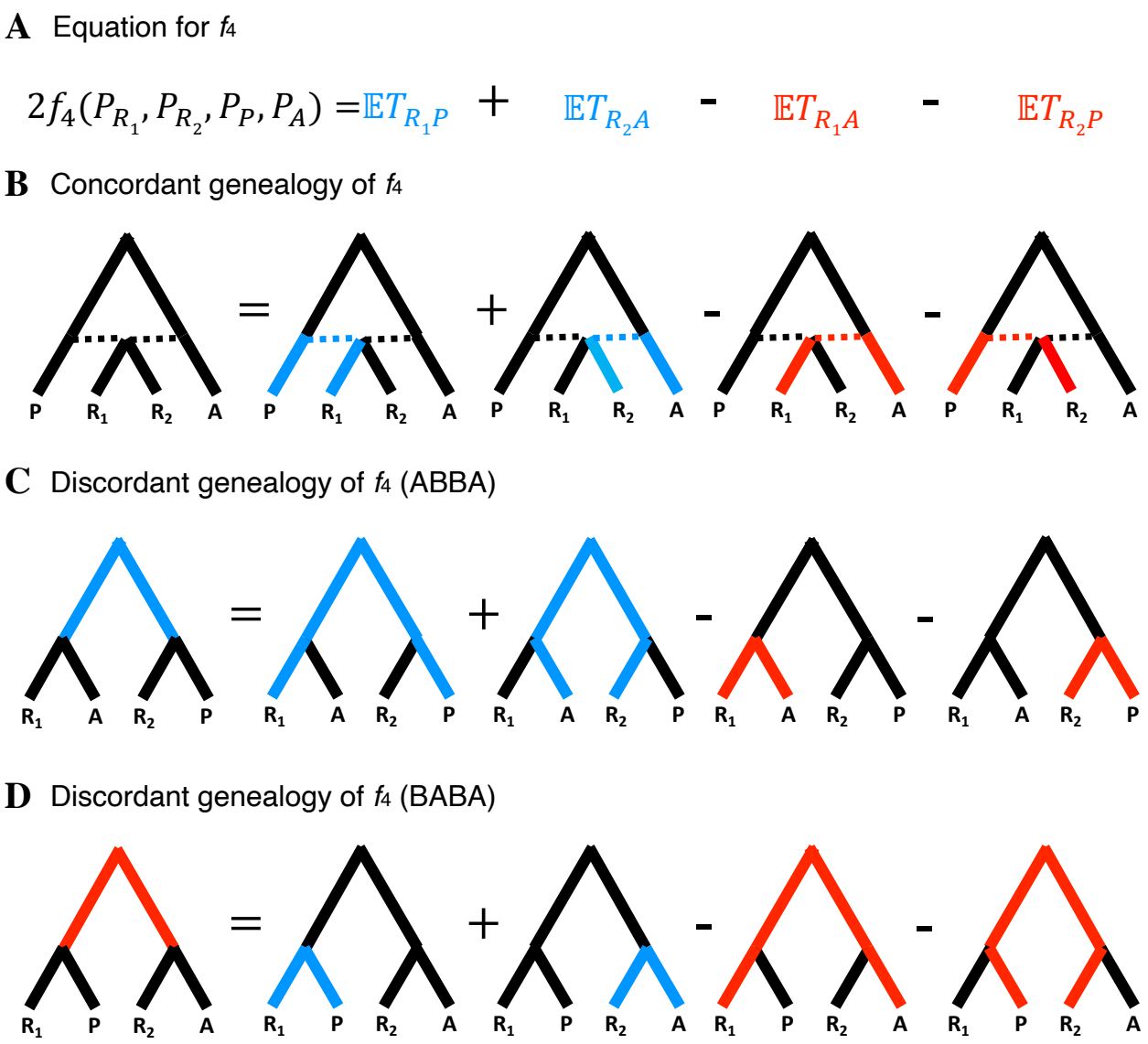


Fig. SM10. Schematic explaining how f_4 behaved in this study. "A" represents the population of *D. abyssinica*. "P" represents the population of *D. praehensilis*. "R1" represents the first populations of *D. rotundata*. "R2" represents the second populations of *D. rotundata*. This figure was adapted from (38).

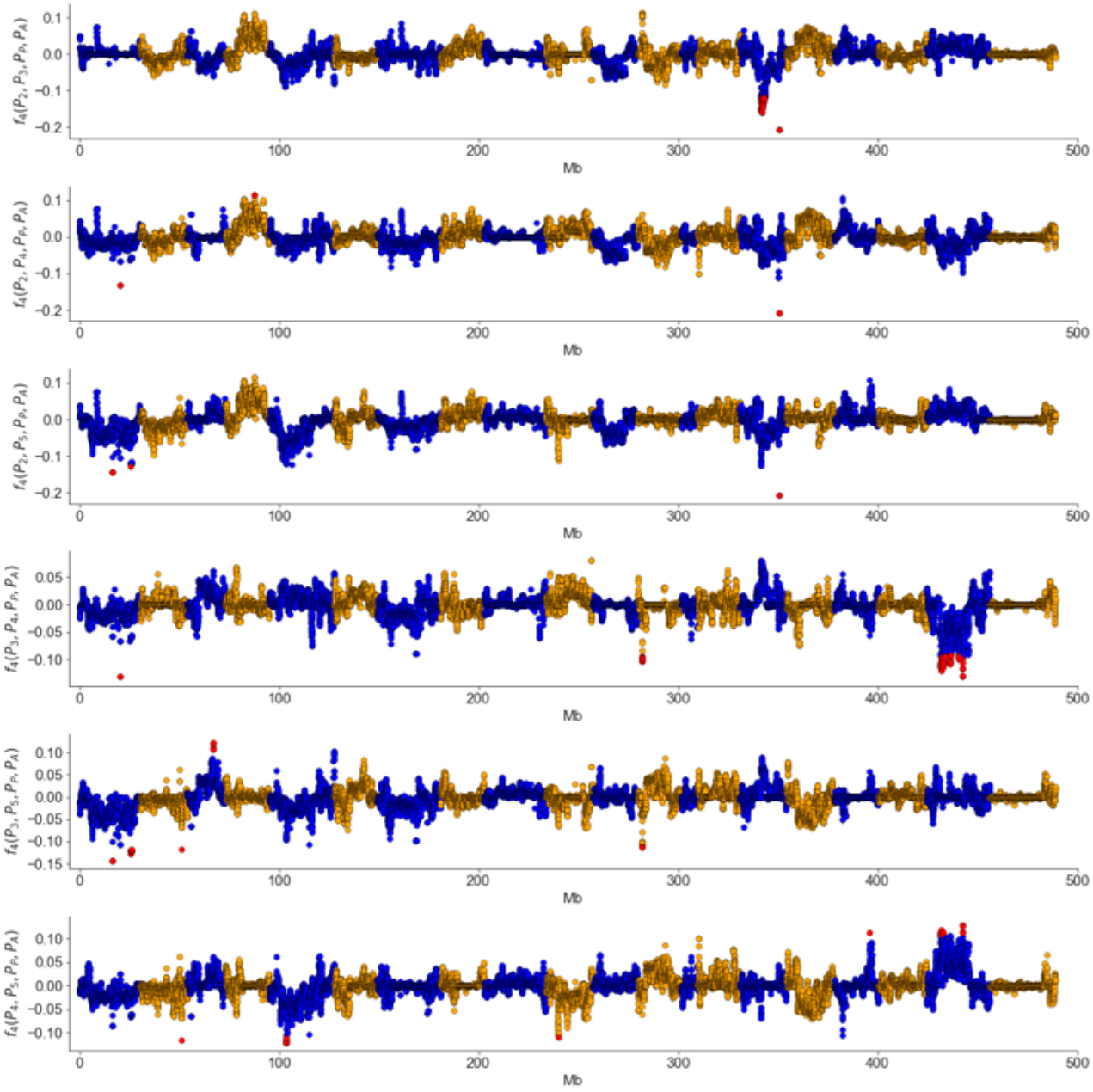


Fig. SM11. f_4 in all possible combinations of clusters excluding cluster 1. Population P*i* represents a population of the cluster *i*.

CHAPTER 3: GENERAL DISCUSSION

Population genomics and cytogenetics studies have revealed important domestication processes in *Dioscorea* species, but many questions still remain. For example, we still do not know the key traits and the genes involved in yam domestications, although some studies have identified genes showing signature of selection in *D. rotundata* including *SWEETIE* gene in our study (Akakpo et al., 2017; Scarcelli et al., 2019; Sugihara et al., 2020). *D. abyssinica* and *D. praehensilis,* the wild relatives of *D. rotundata*, are subjected to an on-going practice of 'ennoblement'. Additionally, it has been shown that the cultivars introduced by 'ennoblement' are indeed hybrids between the cultivated and wild yams (Chaïr et al., 2010; Scarcelli et al., 2006). These findings probably indicate that the wild species cannot directly be domesticated to become cultivars and that hybridization was necessary to generate white Guinea yam cultivars. Similar interspecific hybridization was also reported in *D. alata* (Chaïr et al., 2016). Consequently, analyzing hybridization is important to understand what attributes characterize *D. rotundata* and other cultivated yams. Probably, *D. rotundata* was established as a cultivar as a result of heterosis derived from the hybridization between *D. abyssinica* and *D. praehensilis*.

Understanding the genomes of crop wild relatives would facilitate efficient breeding programs. Crop wild relatives are expected to have potentially beneficial alleles that are not available in the cultivars. The farmers unconsciously introduce these beneficial alleles to the cultivars presumably by 'ennoblement'. Since the genomic regions containing the beneficial alleles should be affected by selective sweeps, population genomics analyses may be able to identify these regions (Akakpo et al., 2017; Scarcelli et al., 2019; Sugihara et al., 2020). Currently, there is no evidence that these candidate selective sweeps affected any phenotypes. However future functional studies of the identified genes would reveal their impact on the change of traits in the crops.

Another standing question is how many times the domestication processes occurred in the various cultivated *Dioscorea* species. A recent study hypothesized multiple domestication processes of *D. alata* in separate regions (Sharif et al., 2020). The cultivated yam landraces from Southern Ethiopian are phylogenetically close to the cultivated gene pools of *D. rotundata*, but they were clearly separate from Nigerian *D. rotundata* (Tamiru et al., 2007). Although the model-based population genetics/genomics is needed to infer the detailed

demographic history, this result may suggest independent domestication processes of *D. rotundata* in Ethiopia (or East Africa) and Nigeria.

The importance of hybridization and polyploidization for the domestication of *Dioscorea* species has been discussed. Some of these events appear to have played an important role in yam domestication. In recent years, our knowledge of yam domestication has dramatically improved thanks to the advances in sequencing technologies and statistical methods for population genomics analysis. These developments also allowed us to identify, among others, the transition of the sex-determination system in the section Enantiophyllum. Future studies should further unravel the complex evolutionary history of *Dioscorea* species including hybridization, polyploidization, and sexual/asexual propagation.

SUPPLEMENTARY DATA

Table S1. All sequence information of Guinea yam accessions

Sample			Fastq size		Aligned bam information				Sequence platform	Cluster	Accession No.
Name	IITA name	Ploidy level	Original (Gbp)	Filtered (Gbp)	Aligned (Gbp)	Unmapped (Gbp)	Coverage (%)	Depth			
TDr04_219	TDr04_219	-	38.26	33.10	25.95	0.32	89.7	49.93	MiSeq, HiSeq4000, GAIIx	-	DRR208404,DRR208405,DRR063085
TDr97_777	TDr97_777	-	50.20	43.48	32.72	0.94	89.8	62.90	MiSeq,HiSeq4000,NextSeq500,GAIIx	-	DRR063127,DRR208406,DRR045130-7,DRR063111
TDr96_F1	TDr96_F1	-	16.77	21.34	17.66	0.04	90.3	33.74	MiSeq	-	DRR027644
DRS_001	TDr2946A	2	12.70	11.32	9.28	0.12	86.4	18.53	HiSeq4000	-	DRR208876
DRS_002	TDr1489A	2	8.09	7.02	5.27	0.05	79.0	11.51	HiSeq4000	-	DRR208761
DRS_003	TDr2284A	2	15.28	13.18	9.50	0.16	87.3	18.78	HiSeq4000,NextSeq500	-	DRR208762,DRR208884
DRS_004	TDr1499A	2	13.65	11.70	8.55	0.13	86.9	16.99	HiSeq4000,NextSeq500	cluster2	DRR208763,DRR208885
DRS_006	TDr1509A	2	13.47	11.10	8.13	0.16	86.6	16.19	HiSeq4000,NextSeq500	-	DRR208764,DRR208886
DRS_007	TDr1510A	2	12.30	10.19	7.81	0.14	87.2	15.46	HiSeq4000,NextSeq500	-	DRR208765,DRR208887
DRS_009	TDr3782A	2	13.23	11.21	7.62	0.14	89.0	14.77	HiSeq4000,NextSeq500	-	DRR208766,DRR208888
DRS_010	TDr1533A	2	13.31	11.18	8.32	0.17	86.6	16.58	HiSeq4000,NextSeq500	cluster3	DRR208767,DRR208889
DRS_011	TDr1543A	2	13.22	11.17	7.33	0.18	86.7	14.59	HiSeq4000,NextSeq500	cluster3	DRR208768,DRR208890
DRS_012	TDr1858C	2	14.22	12.56	9.99	0.11	85.6	20.14	HiSeq4000	-	DRR208877
DRS_013	TDr1576A	2	12.81	11.47	9.91	0.15	87.4	19.57	HiSeq4000,NextSeq500	cluster3	DRR208769,DRR208891
DRS_014	TDr1585A	2	14.01	11.78	7.58	0.15	87.1	15.02	HiSeq4000,NextSeq500	-	DRR208770,DRR208892
DRS_015	TDr1585C	2	15.02	12.78	8.02	0.16	87.0	15.92	HiSeq4000,NextSeq500	cluster2	DRR208771,DRR208893
DRS_016	TDr1598A	3	13.80	11.74	7.54	0.43	86.6	15.02	HiSeq4000,NextSeq500	cluster1	DRR208772,DRR208894
DRS_017	TDr1622A	2	12.71	10.81	7.02	0.15	86.8	13.96	HiSeq4000,NextSeq500	cluster2	DRR208773,DRR208895
DRS_018	TDr1628A	2	8.08	7.00	5.24	0.05	77.7	11.65	HiSeq4000	-	DRR208774
DRS_019	TDr1631C	2	15.29	13.62	11.34	0.17	87.8	22.31	HiSeq4000,NextSeq500	-	DRR208775,DRR208896
DRS_020	TDr1649A	2	12.64	10.83	7.22	0.15	86.4	14.42	HiSeq4000,NextSeq500	cluster3	DRR208776,DRR208897
DRS_021	TDr1650B	2	12.42	10.56	8.18	0.15	87.0	16.23	HiSeq4000,NextSeq500	cluster2	DRR208777,DRR208898
DRS_022	TDr1653A	2	12.68	10.86	6.92	0.13	86.5	13.80	HiSeq4000,NextSeq500	-	DRR208778,DRR208899
DRS_023	TDr1655A	2	12.24	10.36	6.42	0.16	86.0	12.89	HiSeq4000,NextSeq500	cluster5	DRR208779,DRR208900
DRS_024	TDr1663A	2	13.69	11.55	8.68	0.17	86.9	17.22	HiSeq4000,NextSeq500	-	DRR208780,DRR208901
DRS_025	TDr1686A	2	12.58	11.14	8.11	0.12	88.2	15.87	HiSeq4000,NextSeq500	cluster4	DRR208781,DRR208902
DRS_026	TDr1707A	2	13.51	11.93	9.04	0.15	88.7	17.58	HiSeq4000,NextSeq500	-	DRR208782,DRR208903
DRS_027	TDr1709A	2	16.91	15.10	12.04	0.22	86.1	24.13	HiSeq4000	-	DRR208878
DRS_028	TDr1711A	2	7.64	6.65	5.15	0.04	80.3	11.06	HiSeq4000	-	DRR208783
DRS_029	TDr3872A	2	12.61	11.23	8.95	0.15	87.1	17.73	HiSeq4000,NextSeq500	-	DRR208784,DRR208904
DRS_030	TDr1732A	2	13.95	12.38	9.83	0.16	88.2	19.23	HiSeq4000,NextSeq500	cluster4	DRR208785,DRR208905
DRS_031	TDr1735A	2	13.02	11.59	7.48	0.15	86.5	14.93	HiSeq4000,NextSeq500	-	DRR208786,DRR208906
DRS_032	TDr2029A	2	11.77	10.51	7.03	0.14	85.7	14.16	HiSeq4000,NextSeq500	-	DRR208787,DRR208907
DRS_033	TDr1760A	2	11.42	9.96	8.19	0.08	87.9	16.09	HiSeq4000,NextSeq500	cluster4	DRR208788,DRR208908
DRS_034	TDr1763C	2	16.32	14.06	11.22	0.09	87.7	22.07	HiSeq4000,NextSeq500	cluster4	DRR208789,DRR208909
DRS_035	TDr1804A	2	12.63	10.98	8.56	0.10	87.7	16.85	HiSeq4000,NextSeq500	cluster4	DRR208790,DRR208910
DRS_036	TDr1775A	3	11.44	9.96	7.52	0.27	87.5	14.82	HiSeq4000,NextSeq500	cluster1	DRR208791,DRR208911
DRS_037	TDr1798A	2	8.14	7.02	5.25	0.05	79.8	11.35	HiSeq4000	-	DRR208792
DRS_038	TDr1805A	2	12.60	11.14	9.16	0.08	86.7	18.22	HiSeq4000,NextSeq500	-	DRR208793,DRR208912
DRS_039	TDr1807A	3	11.35	10.05	7.67	0.26	87.6	15.11	HiSeq4000,NextSeq500	cluster1	DRR208794,DRR208913
DRS_040	TDr1829A	2	11.26	9.77	7.35	0.10	85.9	14.77	HiSeq4000,NextSeq500	cluster3	DRR208795,DRR208914
DRS_041	TDr1850A	2	11.35	10.03	8.20	0.09	86.5	16.36	HiSeq4000,NextSeq500	cluster5	DRR208796,DRR208915
DRS_042	TDr1899A	2	13.90	12.01	7.89	0.15	86.8	15.68	HiSeq4000,NextSeq500	-	DRR208797,DRR208916
DRS_043	TDr1922C	2	7.43	6.29	4.76	0.05	79.7	10.31	HiSeq4000	-	DRR208798
DRS_044	TDr1935A	2	11.67	10.14	7.65	0.14	86.7	15.23	HiSeq4000,NextSeq500	-	DRR208799,DRR208917
DRS_045	TDr2608A	2	12.53	10.86	7.99	0.18	86.7	15.91	HiSeq4000,NextSeq500	-	DRR208800,DRR208918
DRS_046	TDr2041B	2	14.19	12.52	10.16	0.16	86.6	20.26	HiSeq4000,NextSeq500	-	DRR208801,DRR208919

DRS_047	TDr2121A	2	11.61	10.17	7.98	0.10	86.5	15.93	HiSeq4000,NextSeq500	cluster3	DRR208802,DRR208920
DRS_048	TDr2155A	3	13.17	11.47	7.84	0.41	87.6	15.44	HiSeq4000,NextSeq500	cluster1	DRR208803,DRR208921
DRS_049	TDr2159A	2	11.28	9.93	7.79	0.20	86.9	15.46	HiSeq4000,NextSeq500	-	DRR208804,DRR208922
DRS_050	TDr2161C	3	12.97	11.43	7.79	0.42	87.7	15.33	HiSeq4000,NextSeq500	cluster1	DRR208805,DRR208923
DRS_051	TDr2167A	3	11.86	10.44	7.60	0.36	87.6	14.98	HiSeq4000,NextSeq500	cluster1	DRR208806,DRR208924
DRS_053	TDr2207A	2	11.41	9.81	7.75	0.08	86.2	15.51	HiSeq4000,NextSeq500	-	DRR208807,DRR208925
DRS_054	TDr2210A	2	10.71	9.52	7.15	0.14	85.5	14.42	HiSeq4000	cluster2	DRR208879
DRS_055	TDr3311B	2	13.88	12.29	9.69	0.18	88.2	18.96	HiSeq4000,NextSeq500	cluster4	DRR208808,DRR208926
DRS_056	TDr2262C	2	11.63	10.36	8.70	0.19	86.5	17.36	HiSeq4000,NextSeq500	-	DRR208809,DRR208927
DRS_057	TDr2320A	2	8.95	7.70	6.41	0.08	82.2	13.46	HiSeq4000,NextSeq500	cluster3	DRR208928
DRS_058	TDr2484A	2	12.62	11.21	9.03	0.13	86.8	17.96	HiSeq4000,NextSeq500	-	DRR208810,DRR208929
DRS_059	TDr2973A	2	11.15	9.46	7.31	0.10	83.6	15.11	HiSeq4000,NextSeq500	-	DRR208930
DRS_060	TDr2425B	2	10.38	8.92	7.47	0.07	86.5	14.89	HiSeq4000,NextSeq500	-	DRR208931
DRS_061	TDr2427B	3	12.28	11.17	8.47	0.38	87.9	16.63	HiSeq4000,NextSeq500	cluster1	DRR208811,DRR208932
DRS_062	TDr2435A	2	7.61	6.72	5.79	0.05	83.2	12.01	HiSeq4000	cluster3	DRR208812
DRS_063	TDr2439A	2	10.11	9.03	7.64	0.06	86.7	15.22	HiSeq4000,NextSeq500	-	DRR208813,DRR208933
DRS_064	TDr2453A	2	13.41	12.08	9.89	0.16	87.3	19.56	HiSeq4000,NextSeq500	-	DRR208814,DRR208934
DRS_065	TDr2491A	2	13.74	12.46	10.23	0.15	87.1	20.28	HiSeq4000,NextSeq500	-	DRR208815,DRR208935
DRS_066	TDr3569A	2	15.47	14.08	10.11	0.15	88.6	19.70	HiSeq4000,NextSeq500	cluster4	DRR208816,DRR208936
DRS_067	TDr2533C	2	11.24	10.03	7.83	0.12	85.5	15.80	HiSeq4000	cluster3	DRR208880
DRS_068	TDr2554A	3	16.37	14.91	9.73	0.57	88.7	18.93	HiSeq4000,NextSeq500	cluster1	DRR208817,DRR208937
DRS_069	TDr2575A	2	8.89	8.03	7.04	0.09	87.2	13.94	HiSeq4000,NextSeq500	-	DRR208938
DRS_070	TDr2636B	2	8.88	7.89	6.48	0.09	85.9	13.02	HiSeq4000,NextSeq500	-	DRR208939
DRS_071	TDr2674A	2	8.56	7.72	6.81	0.07	86.7	13.56	HiSeq4000,NextSeq500	cluster5	DRR208940
DRS_072	TDr2713A	2	13.04	11.77	9.86	0.13	86.9	19.57	HiSeq4000,NextSeq500	-	DRR208818,DRR208941
DRS_073	TDr1684A	2	15.55	13.87	11.48	0.08	87.0	22.79	HiSeq4000,NextSeq500	-	DRR208819,DRR208942
DRS_074	TDr2948A	2	11.67	10.50	8.98	0.07	87.6	17.69	HiSeq4000,NextSeq500	-	DRR208820,DRR208943
DRS_075	TDr2965A	2	12.16	10.98	9.64	0.10	88.4	18.83	HiSeq4000,NextSeq500	-	DRR208944
DRS_076	TDr2968A	2	9.19	8.21	6.80	0.09	84.5	13.88	HiSeq4000	-	DRR208881
DRS_077	TDr2975A	2	11.13	9.98	8.47	0.08	87.6	16.69	HiSeq4000,NextSeq500	cluster4	DRR208945
DRS_078	TDr4067A	3	12.84	11.59	8.41	0.43	87.9	16.52	HiSeq4000,NextSeq500	cluster1	DRR208821,DRR208946
DRS_079	TDr2577A	2	13.00	11.79	8.57	0.14	88.8	16.65	HiSeq4000,NextSeq500	cluster2	DRR208822,DRR208947
DRS_080	TDr3325A	2	13.81	12.10	10.14	0.19	87.3	20.05	HiSeq4000,NextSeq500	cluster3	DRR208823,DRR208948
DRS_081	TDr3470A	2	9.57	8.49	6.66	0.11	84.7	13.58	HiSeq4000	-	DRR208882
DRS_082	TDr3436A	2	9.71	6.42	5.15	0.05	80.8	11.01	HiSeq4000,NextSeq500	-	DRR208824,DRR208949
DRS_083	TDr3447B	2	12.45	9.46	6.69	0.11	85.7	13.47	HiSeq4000,NextSeq500	-	DRR208825,DRR208950
DRS_084	TDr3519A	3	16.08	14.55	9.55	0.49	88.7	18.59	HiSeq4000,NextSeq500	cluster1	DRR208826,DRR208951
DRS_085	TDr3527A	2	7.58	6.63	5.61	0.05	82.7	11.71	HiSeq4000	cluster5	DRR208827
DRS_086	TDr2276A	2	15.07	13.13	9.37	0.16	87.8	18.43	HiSeq4000,NextSeq500	-	DRR208828,DRR208952
DRS_087	TDr3576A	2	17.05	13.22	10.14	0.16	85.7	20.42	HiSeq4000,NextSeq500	-	DRR208953
DRS_088	TDr3624B	2	9.68	7.84	6.21	0.09	83.3	12.86	HiSeq4000,NextSeq500	cluster5	DRR208954
DRS_089	TDr2503A	2	10.12	9.07	7.50	0.09	85.6	15.11	HiSeq4000	-	DRR208883
DRS_090	TDr3678A	2	15.57	13.52	9.08	0.16	87.4	17.93	HiSeq4000,NextSeq500	-	DRR208829,DRR208955
DRS_091	TDr3719A	2	10.51	8.83	7.13	0.12	85.5	14.39	HiSeq4000,NextSeq500	-	DRR208956
DRS_092	TDr3828B	2	14.56	13.01	9.32	0.14	88.0	18.29	HiSeq4000,NextSeq500	-	DRR208830,DRR208957
DRS_093	TDr3842A	2	16.92	14.92	12.90	0.15	88.5	25.16	HiSeq4000,NextSeq500	-	DRR208958
DRS_094	TDr3863A	2	12.26	10.95	9.22	0.08	87.0	18.29	HiSeq4000,NextSeq500	-	DRR208831,DRR208959
DRS_095	TDr3955C	2	12.58	11.25	9.49	0.13	86.8	18.86	HiSeq4000,NextSeq500	-	DRR208832,DRR208960
DRS_096	TDr2090B	2	11.97	10.73	8.40	0.14	86.5	16.77	HiSeq4000,NextSeq500	cluster5	DRR208833,DRR208961
DRS_097	TDr1772A	2	11.77	10.45	7.60	0.17	86.4	15.17	HiSeq4000,NextSeq500	-	DRR208834,DRR208962
DRS_098	TDr3357A	2	12.18	10.91	9.10	0.14	87.1	18.02	HiSeq4000,NextSeq500	cluster2	DRR208835,DRR208963

DRS_099	TDr4017A	2	13.05	11.46	8.35	0.21	86.8	16.59	HiSeq4000,NextSeq500	cluster3	DRR208836,DRR208964
DRS_100	TDr3623C	2	13.73	11.91	8.88	0.19	87.8	17.45	HiSeq4000,NextSeq500	cluster4	DRR208837,DRR208965
DRS_101	TDr4100A	2	13.31	11.67	9.37	0.18	87.1	18.57	HiSeq4000,NextSeq500	cluster5	DRR208838,DRR208966
DRS_102	TDr2826A	2	8.63	7.53	6.33	0.05	83.7	13.06	HiSeq4000	-	DRR208839
DRS_103	TDr4155A	2	10.98	9.61	7.67	0.15	87.4	15.14	HiSeq4000,NextSeq500	-	DRR208840,DRR208967
DRS_104	TDr4180A	2	11.13	9.80	7.41	0.17	86.1	14.87	HiSeq4000,NextSeq500	cluster5	DRR208841,DRR208968
DRS_106	TDr2042A	2	11.49	10.05	7.77	0.14	86.4	15.53	HiSeq4000,NextSeq500	-	DRR208842,DRR208969
DRS_165	TDr608	-	10.92	9.99	7.98	0.04	86.1	15.99	HiSeq4000	-	DRR208843
DRS_169	TDrFaketsa	-	13.01	11.98	9.39	0.07	85.2	19.00	HiSeq4000	-	DRR208844
DRS_177	TDrGbangu	-	11.57	10.54	8.17	0.05	84.2	16.74	HiSeq4000	-	DRR208845
DRS_208	TDr09/00362	-	9.01	8.25	6.58	0.04	83.7	13.57	HiSeq4000	-	DRR208846
DRS_211	TDr09/00799	-	10.73	9.84	7.82	0.04	85.3	15.83	HiSeq4000	-	DRR208847
DRS_212	TDrMeccakusa	-	8.41	7.64	6.18	0.03	84.4	12.64	HiSeq4000	-	DRR208848
DRS_213	TDr09/09132	-	9.97	9.14	7.40	0.04	85.3	14.98	HiSeq4000	-	DRR208849
DRS_220	TDrOjuiyawo	-	7.58	6.97	5.80	0.04	85.7	11.69	HiSeq4000	-	DRR208850
DRS_253	TDr2119	-	9.33	8.57	7.03	0.05	84.6	14.34	HiSeq4000	cluster2	DRR208851
DRS_259	TDr2347	-	9.94	9.10	6.98	0.05	85.1	14.16	HiSeq4000	cluster4	DRR208852
DRS_282	TDrOgoja	-	12.34	11.27	8.64	0.06	85.0	17.54	HiSeq4000	-	DRR208853
DRS_293	TDr10/00077	-	10.65	9.72	7.73	0.05	83.1	16.05	HiSeq4000	-	DRR208854
DRS_297	TDrGbongi	-	9.56	8.63	6.51	0.06	84.2	13.35	HiSeq4000	-	DRR208855
DRS_307	TDr10/00125	-	8.97	8.23	6.41	0.04	83.3	13.27	HiSeq4000	-	DRR208856
DRS_312	TDrLagos	-	7.85	7.13	5.63	0.05	82.8	11.73	HiSeq4000	-	DRR208857
DRS_318	TDrHembakwase	-	9.25	8.48	6.86	0.05	85.6	13.84	HiSeq4000	-	DRR208858
DRS_320	TDr89/02157	-	11.44	10.42	8.06	0.05	85.4	16.30	HiSeq4000	-	DRR208859
DRS_322	TDr97/00632	-	8.64	7.92	6.19	0.05	82.3	12.98	HiSeq4000	-	DRR208860
DRS_324	TDr00/02405	-	11.07	10.00	7.70	0.05	84.4	15.74	HiSeq4000	-	DRR208861
DRS_325	TDr10/00013	-	10.25	9.28	7.26	0.06	84.0	14.90	HiSeq4000	-	DRR208862
DRS_326	TDr10/00048	-	8.99	8.28	6.81	0.04	84.6	13.90	HiSeq4000	-	DRR208863
DRS_327	TDr10/00179	-	9.13	8.29	6.55	0.05	83.5	13.53	HiSeq4000	-	DRR208864
DRS_328	TDr10/00344	-	10.16	9.27	7.49	0.04	84.8	15.25	HiSeq4000	-	DRR208865
DRS_329	TDr10/00360	-	11.47	10.35	8.04	0.05	84.6	16.41	HiSeq4000	-	DRR208866
DRS_330	TDr10/00459	-	11.39	10.42	8.27	0.05	84.3	16.93	HiSeq4000	-	DRR208867
DRS_331	TDr10/00021	-	10.88	9.96	8.05	0.05	85.6	16.24	HiSeq4000	-	DRR208868
DRS_332	TDr89/02475	-	7.70	7.05	5.97	0.04	85.5	12.05	HiSeq4000	-	DRR208869
DRS_333	TDr89/02677	-	9.64	8.89	7.41	0.05	85.9	14.89	HiSeq4000	-	DRR208870
DRS_334	TDr96/00629	-	9.43	8.64	6.94	0.04	86.3	13.88	HiSeq4000	-	DRR208871
DRS_335	TDr96/01818	-	10.27	9.39	7.53	0.05	86.3	15.07	HiSeq4000	-	DRR208872
DRS_336	TDr99/02562	-	10.56	9.66	7.89	0.05	85.9	15.84	HiSeq4000	-	DRR208873
DRS_337	TDrAkwuchi	-	9.43	8.65	7.11	0.04	86.0	14.28	HiSeq4000	-	DRR208874
DRS_338	TDrDanacha	-	10.57	9.54	7.64	0.06	84.6	15.59	HiSeq4000	-	DRR208875
TDr_001	TDr1492	-	8.93	7.47	6.00	0.05	81.8	12.66	HiSeq4000	cluster3	DRR208563
TDr_002	TDr2262	-	6.84	5.83	4.90	0.04	82.5	10.24	HiSeq4000	-	DRR208564
TDr_003	TDr1533	-	7.61	6.25	5.00	0.05	78.6	10.98	HiSeq4000	cluster3	DRR208565
TDr_004	TDr1559	-	8.65	7.51	5.75	0.07	84.1	11.81	HiSeq4000	cluster4	DRR208566
TDr_005	TDr1577	-	8.77	7.73	6.22	0.22	81.6	13.14	HiSeq4000	cluster3	DRR208567
TDr_006	TDr1598	-	9.47	8.18	5.96	0.07	82.6	12.44	HiSeq4000	cluster2	DRR208568
TDr_007	TDr1615	-	8.48	7.14	5.27	0.16	81.3	11.17	HiSeq4000	cluster1	DRR208569
TDr_008	TDr1628	-	7.36	6.27	5.35	0.05	84.4	10.94	HiSeq4000	-	DRR208570
TDr_009	TDr1669	-	7.41	6.53	4.90	0.03	81.0	10.44	HiSeq4000	cluster2	DRR208571
TDr_010	TDr1707	-	9.48	8.20	6.20	0.06	84.4	12.68	HiSeq4000	-	DRR208572
TDr_011	TDr1717	-	8.85	7.98	6.13	0.05	82.2	12.88	HiSeq4000	-	DRR208573

TDr_012	TDr1763	-	8.62	7.76	6.09	0.05	82.2	12.79	HiSeq4000	-	DRR208574
TDr_013	TDr1769	-	10.00	8.55	6.64	0.23	85.9	13.34	HiSeq4000	cluster1	DRR208575
TDr_014	TDr1799	-	7.81	6.87	4.65	0.03	80.1	10.02	HiSeq4000	-	DRR208576
TDr_015	TDr1825	-	8.01	6.33	5.14	0.05	82.8	10.71	HiSeq4000	cluster4	DRR208577
TDr_016	TDr1876	-	9.56	8.32	6.48	0.21	85.7	13.06	HiSeq4000	cluster1	DRR208578
TDr_017	TDr1937	-	10.02	8.61	6.82	0.06	82.5	14.27	HiSeq4000	cluster3	DRR208579
TDr_018	TDr1939	-	9.87	8.09	6.57	0.05	83.1	13.64	HiSeq4000	cluster2	DRR208580
TDr_019	TDr1949	-	8.22	7.20	6.01	0.05	82.3	12.60	HiSeq4000	cluster2	DRR208581
TDr_020	TDr2015	-	8.50	7.36	5.69	0.17	84.8	11.58	HiSeq4000	cluster1	DRR208582
TDr_021	TDr2028	-	9.44	8.27	6.56	0.21	86.1	13.16	HiSeq4000	cluster1	DRR208583
TDr_022	TDr2038	-	7.88	6.63	5.50	0.03	83.7	11.35	HiSeq4000	-	DRR208584
TDr_023	TDr2050	-	10.79	8.75	6.75	0.08	81.8	14.23	HiSeq4000	-	DRR208585
TDr_024	TDr2059	-	10.16	8.59	6.95	0.06	85.6	14.03	HiSeq4000	cluster4	DRR208586
TDr_025	TDr2090	-	7.64	6.44	4.68	0.05	80.0	10.10	HiSeq4000	-	DRR208587
TDr_026	TDr2104	-	8.55	7.51	5.58	0.06	82.4	11.69	HiSeq4000	cluster4	DRR208588
TDr_027	TDr2110	-	9.84	8.65	6.78	0.21	85.7	13.66	HiSeq4000	cluster1	DRR208589
TDr_028	TDr2211	-	9.65	8.28	6.96	0.05	84.9	14.16	HiSeq4000	-	DRR208590
TDr_029	TDr2080	-	9.78	8.47	7.43	0.07	86.4	14.86	HiSeq4000	-	DRR208591
TDr_030	TDr2349	-	9.61	8.37	7.24	0.12	87.1	14.36	HiSeq4000	-	DRR208592
TDr_031	TDr2363	-	7.70	6.26	4.78	0.04	79.6	10.37	HiSeq4000	cluster2	DRR208593
TDr_032	TDr2406	-	10.33	9.01	7.83	0.06	85.2	15.86	HiSeq4000	cluster2	DRR208594
TDr_033	TDr2432	-	6.70	5.57	4.63	0.04	81.6	9.80	HiSeq4000	cluster2	DRR208595
TDr_034	TDr2439	-	9.57	7.96	6.11	0.06	82.3	12.80	HiSeq4000	-	DRR208596
TDr_035	TDr2458	-	6.94	5.83	5.02	0.04	84.7	10.24	HiSeq4000	cluster4	DRR208597
TDr_036	TDr2502	-	6.51	5.58	4.40	0.04	80.6	9.43	HiSeq4000	cluster5	DRR208598
TDr_037	TDr2581	-	9.62	8.34	7.10	0.10	86.3	14.20	HiSeq4000	cluster4	DRR208599
TDr_038	TDr2645	-	9.37	8.43	6.38	0.05	82.1	13.41	HiSeq4000	cluster3	DRR208600
TDr_039	TDr2674	-	7.65	6.59	5.29	0.04	82.2	11.10	HiSeq4000	cluster5	DRR208601
TDr_040	TDr2681	-	10.16	8.00	5.62	0.18	82.6	11.74	HiSeq4000	cluster1	DRR208602
TDr_041	TDr2683	-	9.63	6.47	4.97	0.09	78.7	10.89	HiSeq4000	cluster3	DRR208603
TDr_042	TDr2687	-	14.48	12.64	11.02	0.10	85.8	22.16	HiSeq4000	cluster3	DRR208604
TDr_043	TDr2701	-	9.63	7.79	6.41	0.06	84.9	13.03	HiSeq4000	-	DRR208605
TDr_044	TDr2724	-	10.14	6.15	4.76	0.10	81.4	10.10	HiSeq4000	-	DRR208606
TDr_045	TDr2694	-	8.06	7.00	6.09	0.05	84.8	12.40	HiSeq4000	cluster2	DRR208607
TDr_046	TDr2770	-	9.33	7.46	5.42	0.07	82.0	11.40	HiSeq4000	cluster4	DRR208608
TDr_047	TDr2936	-	10.09	8.13	5.54	0.09	80.8	11.83	HiSeq4000	-	DRR208609
TDr_048	TDr2965	-	10.01	8.76	7.15	0.06	82.9	14.90	HiSeq4000	-	DRR208610
TDr_049	TDr2973	-	13.14	11.33	8.88	0.28	86.9	17.64	HiSeq4000	cluster1	DRR208611
TDr_050	TDr3002	-	9.89	7.16	5.32	0.08	79.7	11.52	HiSeq4000	cluster3	DRR208612
TDr_051	TDr09/00064	-	8.52	5.64	4.47	0.07	78.5	9.82	HiSeq4000	-	DRR208613
TDr_052	TDr00/00362	-	8.13	7.27	6.03	0.05	84.4	12.32	HiSeq4000	-	DRR208614
TDr_053	TDr05/00589	-	12.86	11.09	9.63	0.08	85.1	19.53	HiSeq4000	-	DRR208615
TDr_054	TDr05/00632	-	7.87	6.74	5.29	0.06	80.3	11.37	HiSeq4000	-	DRR208616
TDr_055	TDr07/00157	-	8.49	7.10	6.10	0.05	83.8	12.57	HiSeq4000	-	DRR208617
TDr_056	TDr09/01932	-	8.47	7.64	6.48	0.05	84.9	13.16	HiSeq4000	-	DRR208618
TDr_057	TDr08/00092	-	8.19	6.98	5.69	0.08	81.2	12.08	HiSeq4000	-	DRR208619
TDr_058	TDr08/00108	-	9.46	8.61	7.02	0.06	85.2	14.21	HiSeq4000	-	DRR208620
TDr_059	TDr08/00122	-	8.98	8.09	6.63	0.09	85.1	13.44	HiSeq4000	-	DRR208621
TDr_061	TDr07/00732	-	10.10	9.16	7.79	0.06	85.1	15.80	HiSeq4000	-	DRR208622
TDr_062	TDr08/00207	-	8.83	7.27	6.13	0.07	84.7	12.49	HiSeq4000	-	DRR208623
TDr_063	TDr08/00617	-	9.35	8.46	7.02	0.06	84.6	14.32	HiSeq4000	-	DRR208624

TDr_064	TDr08/00799	-	11.50	10.44	8.23	0.55	85.7	16.58	HiSeq4000	-	DRR208625
TDr_065	TDr09/00325	-	14.08	12.80	10.42	0.09	85.9	20.96	HiSeq4000	-	DRR208626
TDr_066	TDr96/02433	-	14.54	13.18	10.40	0.15	85.7	20.95	HiSeq4000	-	DRR208627
TDr_067	TDr08/01344	-	15.31	14.04	11.58	0.10	86.3	23.16	HiSeq4000	-	DRR208628
TDr_068	TDr08/01024	-	6.51	5.79	4.89	0.04	84.5	9.99	HiSeq4000	-	DRR208629
TDr_069	TDr09/00023	-	7.32	6.64	5.55	0.05	83.3	11.51	HiSeq4000	-	DRR208630
TDr_070	TDr09/00028	-	9.46	8.59	6.94	0.08	83.9	14.26	HiSeq4000	-	DRR208631
TDr_071	TDr09/00056	-	6.89	5.97	5.12	0.07	84.2	10.49	HiSeq4000	-	DRR208632
TDr_072	TDr09/00070	-	8.13	7.31	6.06	0.05	84.1	12.44	HiSeq4000	-	DRR208633
TDr_073	TDr09/00091	-	7.81	7.01	5.93	0.05	83.7	12.23	HiSeq4000	-	DRR208634
TDr_074	TDr09/00104	-	8.88	8.12	6.80	0.05	85.6	13.71	HiSeq4000	-	DRR208635
TDr_075	TDr09/00108	-	8.73	7.55	6.06	0.06	83.7	12.50	HiSeq4000	-	DRR208636
TDr_076	TDr09/00114	-	7.66	6.38	5.28	0.04	82.9	11.01	HiSeq4000	-	DRR208637
TDr_077	TDr09/00125	-	8.29	7.12	5.83	0.05	82.8	12.16	HiSeq4000	-	DRR208638
TDr_078	TDr09/00134	-	5.53	4.51	3.74	0.03	79.3	8.14	HiSeq4000	-	DRR208639
TDr_079	TDr09/00248	-	9.28	8.19	6.26	0.04	82.5	13.09	HiSeq4000	-	DRR208640
TDr_080	TDr09/00350	-	8.54	7.59	6.36	0.03	83.2	13.18	HiSeq4000	-	DRR208641
TDr_081	TDr99/02789	-	5.88	4.71	3.79	0.02	78.9	8.29	HiSeq4000	-	DRR208642
TDr_082	TDr11/00263.1	-	5.02	4.25	3.73	0.04	82.3	7.81	HiSeq4000	-	DRR208643
TDr_083	TDr08/00161	-	7.23	6.40	5.24	0.06	82.7	10.93	HiSeq4000	-	DRR208644
TDr_084	TDr11/00799	-	13.32	11.78	9.92	0.07	88.2	19.40	HiSeq4000	-	DRR208645
TDr_085	TDr11/01041	-	8.72	7.87	6.55	0.06	86.2	13.12	HiSeq4000	-	DRR208646
TDr_086	TDr12/00474	-	8.47	7.56	5.81	0.06	83.1	12.07	HiSeq4000	-	DRR208647
TDr_087	TDr08/00146	-	8.96	7.98	6.64	0.04	86.6	13.23	HiSeq4000	-	DRR208648
TDr_088	TDrAlumaco	-	10.90	9.16	7.27	0.08	82.2	15.26	HiSeq4000	-	DRR208649
TDr_089	TDrHembakoase	-	6.40	5.79	5.03	0.04	84.8	10.23	HiSeq4000	-	DRR208650
TDr_090	TDr89/02665	-	11.01	9.62	8.24	0.08	86.3	16.48	HiSeq4000	-	DRR208651
TDr_091	TDr05/00046	-	8.32	7.32	5.35	0.06	80.9	11.41	HiSeq4000	-	DRR208652
TDr_092	TDr05/00432	-	8.65	7.33	6.39	0.08	83.8	13.16	HiSeq4000	-	DRR208653
TDr_093	TDr05/00389	-	5.53	4.60	3.74	0.03	79.4	8.14	HiSeq4000	-	DRR208654
TDr_094	TDr08/00023	-	7.24	6.10	5.23	0.05	82.7	10.93	HiSeq4000	-	DRR208655
TDr_095	TDr08/00115	-	7.82	6.64	5.84	0.08	85.6	11.78	HiSeq4000	-	DRR208656
TDr_096	TDr08/00197	-	9.05	7.81	6.66	0.04	85.4	13.45	HiSeq4000	-	DRR208657
TDr_097	TDr08/00974	-	6.68	5.61	4.93	0.04	85.2	9.99	HiSeq4000	-	DRR208658
TDr_098	TDr08/00896	-	8.32	7.54	6.41	0.04	85.1	12.99	HiSeq4000	-	DRR208659
TDr_099	TDr08/00841	-	11.21	9.83	7.43	0.09	85.0	15.09	HiSeq4000	-	DRR208660
TDr_100	TDr0836	-	7.49	6.51	4.97	0.06	79.3	10.82	HiSeq4000	-	DRR208661
TDr_101	TDr1686	-	10.98	9.61	8.03	0.18	86.9	15.96	HiSeq4000	cluster4	DRR208662
TDr_102	TDr3010	-	12.57	11.02	9.10	0.20	85.5	18.36	HiSeq4000	-	DRR208663
TDr_103	TDr3357	-	11.48	10.24	8.63	0.13	85.5	17.41	HiSeq4000	cluster2	DRR208664
TDr_104	TDr3408	-	11.17	9.98	8.39	0.13	86.1	16.81	HiSeq4000	-	DRR208665
TDr_105	TDr3430	-	9.58	8.42	7.19	0.14	86.3	14.38	HiSeq4000	-	DRR208666
TDr_106	TDr3519	-	9.88	8.82	6.71	0.29	85.9	13.48	HiSeq4000	cluster1	DRR208667
TDr_107	TDr3567	-	10.00	8.74	7.24	0.21	85.4	14.63	HiSeq4000	-	DRR208668
TDr_108	TDr3569	-	10.02	8.88	7.43	0.15	86.6	14.81	HiSeq4000	cluster4	DRR208669
TDr_109	TDr3579	-	8.82	7.71	5.91	0.30	85.8	11.88	HiSeq4000	cluster1	DRR208670
TDr_110	TDr3592	-	8.67	7.71	5.90	0.25	85.7	11.89	HiSeq4000	cluster1	DRR208671
TDr_111	TDr3610	-	10.20	9.02	7.46	0.17	86.6	14.86	HiSeq4000	cluster4	DRR208672
TDr_112	TDr3663	-	10.92	9.65	8.10	0.15	86.3	16.21	HiSeq4000	-	DRR208673
TDr_113	TDr3814	-	11.07	9.68	8.13	0.13	86.7	16.19	HiSeq4000	-	DRR208674
TDr_114	TDr3881	-	11.82	10.57	9.07	0.14	86.6	18.07	HiSeq4000	-	DRR208675

TDr_115	TDr4028	-	11.29	9.98	8.31	0.15	86.9	16.51	HiSeq4000	cluster4	DRR208676
TDr_116	TDr08/00641	-	11.86	10.49	8.78	0.18	85.5	17.72	HiSeq4000	-	DRR208677
TDr_117	TDr08/00756	-	9.84	8.60	7.29	0.18	85.4	14.73	HiSeq4000	-	DRR208678
TDr_118	TDr09/00131	-	9.12	7.95	6.77	0.12	84.9	13.76	HiSeq4000	-	DRR208679
TDr_119	TDr1569	-	8.97	7.66	6.44	0.10	85.2	13.05	HiSeq4000	-	DRR208680
TDr_120	TDr2931	-	9.01	7.86	6.83	0.10	85.1	13.84	HiSeq4000	cluster2	DRR208681
TDr_121	TDr2331.1	-	10.29	9.05	7.64	0.11	84.7	15.57	HiSeq4000	-	DRR208682
TDr_122	TDr1958	-	8.66	7.55	5.85	0.22	85.6	11.80	HiSeq4000	cluster1	DRR208683
TDr_123	TDr1905	-	11.62	10.34	8.58	0.18	85.2	17.37	HiSeq4000	cluster5	DRR208684
TDr_124	TDr1928	-	10.93	9.73	8.29	0.11	86.1	16.61	HiSeq4000	-	DRR208685
TDr_125	TDr3322	-	9.48	8.20	6.69	0.17	86.0	13.43	HiSeq4000	cluster4	DRR208686
TDr_126	TDr2048	-	9.82	8.67	7.23	0.15	85.5	14.59	HiSeq4000	-	DRR208687
TDr_127	TDr2126	-	10.08	8.85	7.33	0.14	85.8	14.75	HiSeq4000	-	DRR208688
TDr_128	TDr2249	-	9.38	8.29	6.23	0.23	85.5	12.59	HiSeq4000	cluster1	DRR208689
TDr_129	TDr2297	-	9.97	8.72	6.53	0.30	85.7	13.16	HiSeq4000	cluster1	DRR208690
TDr_130	TDr2342	-	11.08	9.61	7.73	0.16	86.2	15.48	HiSeq4000	cluster4	DRR208691
TDr_131	TDr2355	-	10.27	9.06	6.82	0.29	85.8	13.72	HiSeq4000	cluster1	DRR208692
TDr_132	TDr2564	-	8.75	7.63	6.49	0.13	85.7	13.07	HiSeq4000	-	DRR208693
TDr_133	TDr2698	-	10.55	9.12	7.30	0.16	85.7	14.72	HiSeq4000	cluster4	DRR208694
TDr_134	TDr2974	-	12.48	11.01	8.18	0.32	86.5	16.33	HiSeq4000	cluster1	DRR208695
TDr_135	TDr2975	-	9.23	8.17	6.20	0.24	85.6	12.50	HiSeq4000	cluster1	DRR208696
TDr_136	TDr3507	-	10.29	9.05	7.67	0.11	85.3	15.52	HiSeq4000	cluster2	DRR208697
TDr_137	TDr3006	-	10.03	8.85	7.68	0.12	85.1	15.59	HiSeq4000	cluster5	DRR208698
TDr_138	TDr08/00091	-	7.14	6.30	5.43	0.08	83.8	11.18	HiSeq4000	-	DRR208699
TDr_139	TDr08/01464	-	7.29	6.42	5.59	0.06	84.4	11.44	HiSeq4000	-	DRR208700
TDr_140	TDr08/00989	-	6.96	6.07	5.18	0.09	83.3	10.72	HiSeq4000	-	DRR208701
TDr_141	TDr09/00050	-	7.43	6.50	5.51	0.07	84.3	11.29	HiSeq4000	-	DRR208702
TDr_142	TDr09/00055	-	9.41	8.24	6.96	0.12	84.1	14.28	HiSeq4000	-	DRR208703
TDr_144	TDr09/00061	-	9.15	7.96	6.85	0.09	85.6	13.80	HiSeq4000	-	DRR208704
TDr_145	TDr09/00123	-	8.34	7.28	6.13	0.10	83.9	12.62	HiSeq4000	-	DRR208705
TDr_146	TDr09/00124	-	8.76	7.72	6.59	0.12	85.0	13.39	HiSeq4000	-	DRR208706
TDr_147	TDr09/00220	-	13.21	11.50	9.30	0.15	85.8	18.70	HiSeq4000	-	DRR208707
TDr_148	TDr09/00280.1	-	8.31	7.30	6.22	0.08	84.4	12.73	HiSeq4000	-	DRR208708
TDr_149	TDr09/00324	-	7.26	6.27	5.39	0.09	83.0	11.20	HiSeq4000	-	DRR208709
TDr_150	TDr08/01046	-	12.03	10.56	8.82	0.17	86.2	17.65	HiSeq4000	-	DRR208710
TDr_151	TDrAme	-	14.49	12.87	10.65	0.26	85.9	21.39	HiSeq4000	-	DRR208711
TDr_152	TDrUfenyi	-	12.72	11.18	9.11	0.33	86.6	18.15	HiSeq4000	-	DRR208712
TDr_153	TDr2365	-	12.77	11.26	9.52	0.21	85.4	19.25	HiSeq4000	cluster3	DRR208713
TDr_154	TDr1956	-	10.78	9.74	8.42	0.14	86.0	16.91	HiSeq4000	-	DRR208714
TDr_155	TDr2859	-	11.25	10.13	7.87	0.26	86.3	15.73	HiSeq4000	cluster1	DRR208715
TDr_156	TDr07/000732	-	9.98	8.91	7.72	0.06	85.1	15.66	HiSeq4000	-	DRR208716
TDr_157	TDr08/00764	-	10.88	9.73	8.33	0.09	85.7	16.77	HiSeq4000	-	DRR208717
TDr_158	TDr09/00155	-	9.60	8.63	7.45	0.08	86.5	14.86	HiSeq4000	-	DRR208718
TDr_159	TDr96/01724	-	8.68	7.74	6.72	0.08	84.9	13.65	HiSeq4000	-	DRR208719
TDr_160	TDr08/01287	-	8.54	7.67	6.68	0.06	85.6	13.47	HiSeq4000	-	DRR208720
TDr_161	TDr08/01090	-	9.38	8.41	7.29	0.05	84.9	14.82	HiSeq4000	cluster5	DRR208721
TDr_162	TDr2366	-	9.63	8.50	7.08	0.06	83.8	14.58	HiSeq4000	-	DRR208722
TDr_163	TDr2467	-	9.49	8.55	7.44	0.04	85.2	15.08	HiSeq4000	cluster2	DRR208723
TDr_164	TDr3003	-	9.10	8.14	7.08	0.09	85.2	14.35	HiSeq4000	-	DRR208724
TDr_165	TDr3294	-	8.55	7.66	6.69	0.06	85.9	13.44	HiSeq4000	-	DRR208725
TDr_166	TDr3338	-	11.19	10.03	8.66	0.11	85.2	17.54	HiSeq4000	cluster5	DRR208726

TDr_167	TDr3327	-	10.46	9.35	8.01	0.09	85.3	16.20	HiSeq4000	cluster2	DRR208727
TDr_168	TDr3647	-	10.31	9.19	7.90	0.09	87.4	15.61	HiSeq4000	-	DRR208728
TDr_169	TDr3965	-	11.68	10.48	9.05	0.08	85.5	18.26	HiSeq4000	cluster2	DRR208729
TDr_170	TDr3643	-	10.61	9.47	7.98	0.11	85.0	16.19	HiSeq4000	cluster5	DRR208730
TDr_171	TDr2630	-	8.74	7.71	6.49	0.06	85.7	13.08	HiSeq4000	-	DRR208731
TDr_172	TDr2984	-	11.08	9.72	7.93	0.13	83.2	16.44	HiSeq4000	-	DRR208732
TDr_173	TDr3682	-	9.72	8.74	7.58	0.05	85.1	15.38	HiSeq4000	cluster2	DRR208733
TDr_174	TDr3447	-	8.74	7.75	6.54	0.05	85.1	13.27	HiSeq4000	-	DRR208734
TDr_175	TDr4100	-	8.01	7.17	6.20	0.05	84.6	12.66	HiSeq4000	cluster5	DRR208735
TDr_176	TDr2009	-	10.11	9.05	7.88	0.07	85.2	15.95	HiSeq4000	cluster3	DRR208736
TDr_177	TDr2331.2	-	9.39	8.36	7.18	0.05	84.5	14.66	HiSeq4000	-	DRR208737
TDr_178	TDr3882	-	10.82	9.66	8.22	0.06	85.2	16.65	HiSeq4000	-	DRR208738
TDr_179	TDr2032	-	11.31	10.02	8.45	0.08	85.3	17.09	HiSeq4000	-	DRR208739
TDr_180	TDr11/01036	-	10.84	9.62	8.16	0.07	86.9	16.22	HiSeq4000	-	DRR208740
TDr_181	TDr09/00082	-	9.89	8.79	7.55	0.06	85.5	15.24	HiSeq4000	-	DRR208741
TDr_182	TDr09/00043	-	9.02	7.98	6.81	0.04	85.6	13.74	HiSeq4000	-	DRR208742
TDr_183	TDr09/00364	-	9.62	8.61	7.45	0.07	85.0	15.12	HiSeq4000	-	DRR208743
TDr_184	TDr08/00083	-	9.46	8.39	7.06	0.10	85.0	14.33	HiSeq4000	-	DRR208744
TDr_185	TDr08/01919	-	8.24	7.36	6.40	0.06	85.7	12.89	HiSeq4000	-	DRR208745
TDr_186	TDr09/00216	-	8.75	7.80	6.69	0.05	84.9	13.61	HiSeq4000	-	DRR208746
TDr_187	TDr11/00271	-	8.66	7.68	6.52	0.04	86.5	13.01	HiSeq4000	-	DRR208747
TDr_188	TDr95/18544	-	8.96	8.02	6.90	0.08	85.8	13.87	HiSeq4000	-	DRR208748
TDr_189	TDr11/00263.2	-	9.07	8.09	6.90	0.04	86.1	13.83	HiSeq4000	-	DRR208749
TDr_190	TDr11/00787	-	10.80	9.63	8.38	0.07	87.3	16.56	HiSeq4000	-	DRR208750
TDr_191	TDr09/00385	-	10.35	9.20	7.93	0.05	85.8	15.94	HiSeq4000	-	DRR208751
TDr_192	TDr08/00001	-	11.28	9.96	8.37	0.15	86.8	16.65	HiSeq4000	-	DRR208752
TDr_193	TDr09/00107	-	10.73	9.54	8.19	0.06	85.0	16.63	HiSeq4000	-	DRR208753
TDr_194	TDr08/00882	-	10.05	8.98	7.78	0.07	85.9	15.65	HiSeq4000	-	DRR208754
TDr_195	TDr08/010161	-	8.65	7.60	6.51	0.08	85.9	13.08	HiSeq4000	-	DRR208755
TDr_196	TDr08/01051	-	9.79	8.74	7.58	0.05	85.0	15.38	HiSeq4000	-	DRR208756
TDr_197	TDr08/00292	-	10.60	9.48	8.21	0.10	88.3	16.04	HiSeq4000	-	DRR208757
TDr_198	TDr94/01108	-	10.10	9.00	7.72	0.08	85.4	15.60	HiSeq4000	-	DRR208758
TDr_199	TDr09/00280.1	-	11.04	9.73	8.31	0.09	85.2	16.83	HiSeq4000	-	DRR208708
TDr_200	TDr87/00211	-	9.64	8.60	7.46	0.06	86.5	14.88	HiSeq4000	-	DRR208760

Table S2. For each Figure and Table, the number of SNPs studied, use of triploid *D. rotundata* samples (cluster 1), and use of Scarcelli's samples are indicated.

Figures/tables	No. analyzed SNPs	Triploid *D. rotundata* (in cluster 1)	Dioploid *D. rotundata* (not in cluster 1)	Samples in Scarcelli *et al.* 2019
Fig. 2.2A and B, Fig. 2.3 and 2.4, Table 2.1	6,124,093	yes	yes	-
Table 2.2	5,229,368	no	yes	-
Fig. 2.2C	463,293	no	yes	C/A/P/R
Fig. 2.2D	17,532	no	no	C/A/P
Table 2.4	87,671	no	no	C/A/P
Fig. 2.6B	144	no	yes	A/P/R
Fig. 2.6C and 2.8C	15,461	no	yes	A/P/R
Fig. 2.6D (A vs. R)	649,679	no	yes	A/R
Fig. 2.6D (P vs. R)	579,405	no	yes	P/R
Fig. 2.7 (A vs. P)	362,125	no	no	A/P
Fig. 2.8A	250	yes	yes	C/A/P/R
Fig. 2.9B	2,343,307	no	yes	A/P
Fig. 2.9C	458	no	yes	A/P

C: Cameroonian *D. praehensilis*

A: *D. abyssinica*

P: (Western) *D. praehensilis*

R: *D. rotundata*

Table S3. List of genes in the five outlier loci (chromosome 14, 15, 17, and 19) showing extreme f_4 (P_{25}, P_4, P_P, P_A) values ($|Z(f_4) > 5|$) in Fig. 2.9.

Chromosome	Start	End	GeneID	Annotation
chrom_14	468088	469472	DRNTG_17186.1	(TrEMBL)Predicted protein(HORVV:F2DKZ3)
chrom_14	484029	484961	DRNTG_28166.1	(TrEMBL)Uncharacterized protein(ENSVE:A0A444CGI1)
chrom_14	485725	490867	DRNTG_28165.1	(TrEMBL)Endoplasmic reticulum metallopeptidase 1(ANACO:A0A199W086)
chrom_14	492377	496008	DRNTG_28164.1	Auxin response factor 18(ORYSJ:Q653H7)
chrom_14	496093	496525	DRNTG_28163.1	-
chrom_14	501391	506132	DRNTG_28162.1	Protein ENHANCED DISEASE RESISTANCE 2(ARATH:F4JSE7)
chrom_14	507961	513788	DRNTG_28161.1	Clathrin interactor EPSIN 2(ARATH:Q67YI9)
chrom_14	514348	516233	DRNTG_28160.1	Mitochondrial import inner membrane translocase subunit PAM16 like 2(ARATH:Q93VV9)
chrom_14	516747	519058	DRNTG_28159.1	Cytokinin riboside 5'-monophosphate phosphoribohydrolase LOG4(ARATH:Q9LFH3)
chrom_14	520890	523855	DRNTG_28157.1	Protein CNGC15c(MEDTR:A0A072VMJ3)
chrom_14	521076	521734	DRNTG_28158.1	Protein CNGC15c(MEDTR:A0A072VMJ3)
chrom_14	527056	529173	DRNTG_28156.1	Phytochrome-associated serine/threonine-protein phosphatase(PEA:Q8LSN3)
chrom_14	531504	532632	DRNTG_11714.1	-
chrom_14	544864	552211	DRNTG_11716.1	Phenylalanine ammonia-lyase 3(PETCR:P45729)
chrom_14	550860	554840	DRNTG_11717.1	Phenylalanine ammonia-lyase 3(PETCR:P45729)
chrom_14	565849	567237	DRNTG_11718.1	(TrEMBL)Uncharacterized protein(SETIT:K3ZZF5)
chrom_14	581692	585798	DRNTG_11720.1	E3 ubiquitin-protein ligase WAV3(ARATH:Q9LTA6)
chrom_14	586418	589346	DRNTG_11721.1	General transcription factor IIH subunit 2(ARATH:Q9ZVN9)
chrom_14	589956	591825	DRNTG_11722.1	Probable mannitol dehydrogenase(FRAAN:Q9ZRF1)
chrom_14	607695	608572	DRNTG_25842.1	-
chrom_14	612414	613708	DRNTG_25841.1	(TrEMBL)Uncharacterized protein(MUSAM:M0SPY3)
chrom_14	624755	628872	DRNTG_25840.1	Phenylalanine ammonia-lyase 3(PETCR:P45729)
chrom_14	632310	633887	DRNTG_25839.1	Phenylalanine ammonia-lyase(BROFI:Q42609)
chrom_14	648000	649003	DRNTG_25837.1	-
chrom_14	681229	685056	DRNTG_20830.1	-
chrom_14	695533	696149	DRNTG_12014.1	Ferredoxin--NADP reductase, embryo isozyme, chloroplastic(ORYSJ:O23877)
chrom_14	696349	702726	DRNTG_12015.1	Serine--tRNA ligase, cytoplasmic(ARATH:Q39230)
chrom_14	538878	542983	DRNTG_11715.1	Adenosine kinase 2(ARATH:Q9LZG0)
chrom_14	571194	575569	DRNTG_11719.1	Synaptotagmin-3(ARATH:Q7XA06)
chrom_14	604820	608527	DRNTG_25843.1	Synaptotagmin-3(ARATH:Q7XA06)
chrom_14	637226	641558	DRNTG_25838.1	Adenosine kinase 2(ARATH:Q9LZG0)
chrom_14	717990	726467	DRNTG_12016.1	Adenosine kinase 1(ARATH:Q9SF85)
chrom_15	19356271	19357116	DRNTG_00821.1	-
chrom_15	19362481	19363591	DRNTG_00822.1	-
chrom_15	19603544	19604222	DRNTG_00824.1	ADP,ATP carrier protein, mitochondrial (Fragment)(SOLTU:P27081)
chrom_15	19367597	19459040	DRNTG_00823.1	-
chrom_17	3877215	3877734	DRNTG_07493.1	-
chrom_17	3884570	3885375	DRNTG_07491.1	(TrEMBL)Acyl-coenzyme A thioesterase 13 (Fragment)(9ARAE:A0A1D1Y7U4)
chrom_17	3896927	3897383	DRNTG_07490.1	-
chrom_17	3918976	3920856	DRNTG_07489.1	Cytochrome c-type biogenesis CcmH-like mitochondrial protein(ORYSJ:Q6K7S7)
chrom_17	3924485	3925384	DRNTG_07488.1	Cytochrome c-type biogenesis CcmH-like mitochondrial protein(ORYSI:B8AFK5)
chrom_17	3959026	3959941	DRNTG_07487.1	-
chrom_17	3967202	3969537	DRNTG_07486.1	50S ribosomal protein L1, chloroplastic(SPIOL:Q9LE95)
chrom_17	3969570	3969827	DRNTG_07485.1	-
chrom_17	3978228	3979444	DRNTG_07484.1	24-methylenesterol C-methyltransferase 2(ORYSJ:O82427)
chrom_17	3986569	3989639	DRNTG_07482.1	Cytochrome c-type biogenesis CcmH-like mitochondrial protein(ORYSJ:Q6K7S7)
chrom_17	3986754	3989091	DRNTG_07483.1	Cytochrome c-type biogenesis CcmH-like mitochondrial protein(ORYSJ:Q6K7S7)
chrom_17	4026863	4027853	DRNTG_07481.1	Cytochrome c-type biogenesis CcmH-like mitochondrial protein(ORYSI:B8AFK5)
chrom_17	4106911	4108859	DRNTG_01733.1	Non-specific lipid-transfer protein Cw18(HORVU:Q43871)
chrom_17	4108922	4112102	DRNTG_01734.1	Mitochondrial arginine transporter BAC2(ARATH:Q9CA93)
chrom_17	3875895	3883658	DRNTG_07492.1	Putative E3 ubiquitin-protein ligase LIN-1(LOTJA:C6L7U1)
chrom_17	4095647	4096298	DRNTG_01732.1	-
chrom_17	4053722	4106059	DRNTG_01731.1	Protein SWEETIE(ARATH:F4HRS2)
chrom_19	8230520	8231387	DRNTG_01547.1	-
chrom_19	8307448	8308110	DRNTG_01549.1	-
chrom_19	8314683	8319901	DRNTG_01550.1	-
chrom_19	8319680	8322207	DRNTG_01551.1	Glycerol-3-phosphate acyltransferase RAM2(MEDTR:K7PEY4)
chrom_19	8306157	8311914	DRNTG_01548.1	EID1-like F-box protein 3(ARATH:Q93ZT5)
chrom_19	17790629	17791141	DRNTG_03384.1	Mannose-specific lectin(GALNI:P30617)
chrom_19	17801425	17802462	DRNTG_03385.1	Inorganic phosphate transporter 1-11(ORYSJ:Q94DB8)
chrom_19	17850805	17857145	DRNTG_03386.1	(TrEMBL)uncharacterized protein LOC103722397 isoform X1(PHODC:A0A2H3ZB91)
chrom_19	17964831	17971340	DRNTG_03389.1	Remorin 4.1(ORYSJ:Q7XII4)
chrom_19	17858513	17859406	DRNTG_03387.1	-
chrom_19	17914955	17927446	DRNTG_03388.1	Auxin response factor 12(ORYSI:Q258Y5)

Table S4. Summary of sequence alignment of mapping population.

Sample		Fastq size		Aligned bam information				Sequence platform	Comment		Accession No.
Name	IITA name	Original (Gbp)	Filtered (Gbp)	Aligned (Gbp)	Unmapped (Gbp)	Coverage (%)	Depth		MP2 family Mono parent	MP2 family Male parent	
TDr04_219	TDr04_219	38.26	33.10	17.15	0.32	82.8	35.73	MiSeq, HiSeq4000, GAIIx			DRR208404,DRR208405,DRR063085
TDr97_777	TDr97_777	25.47	22.71	11.20	0.29	79.4	24.35	MiSeq,HiSeq4000,NextSeq500,GAIIx			DRR063127,DRR208406,DRR045130-7,DRR063111
MP2_001	MP2_001	8.20	7.14	4.20	1.00	76.9	9.43	HiSeq4000	-		DRR208407
MP2_002	MP2_002	6.42	5.61	3.45	0.64	73.2	8.13	HiSeq4000	-		DRR208408
MP2_003	MP2_003	5.95	5.11	2.92	0.87	71.6	7.03	HiSeq4000	-		DRR208409
MP2_004	MP2_004	7.13	6.24	3.90	0.70	74.8	8.99	HiSeq4000	-		DRR208410
MP2_005	MP2_005	9.75	8.49	4.59	1.56	75.2	10.53	HiSeq4000	-		DRR208411
MP2_006	MP2_006	7.90	7.01	4.39	0.76	77.2	9.80	HiSeq4000	-		DRR208412
MP2_007	MP2_007	7.50	6.57	4.11	0.75	75.8	9.35	HiSeq4000	-		DRR208413
MP2_008	MP2_008	7.52	6.60	3.93	0.81	74.3	9.13	HiSeq4000	-		DRR208414
MP2_009	MP2_009	7.36	6.48	4.12	0.62	76.3	9.33	HiSeq4000	-		DRR208415
MP2_010	MP2_010	6.49	5.72	3.66	0.55	75.2	8.39	HiSeq4000	-		DRR208416
MP2_011	MP2_011	5.98	5.28	3.41	0.49	77.1	7.63	HiSeq4000	-		DRR208417
MP2_012	MP2_012	8.25	7.31	4.69	0.77	76.9	10.53	HiSeq4000	-		DRR208418
MP2_013	MP2_013	9.33	8.05	4.81	1.00	76.2	10.89	HiSeq4000	-		DRR208419
MP2_014	MP2_014	9.84	8.65	5.56	0.81	78.0	12.32	HiSeq4000	-		DRR208420
MP2_015	MP2_015	11.21	9.80	6.29	0.93	78.5	13.82	HiSeq4000	-		DRR208421
MP2_016	MP2_016	12.97	11.36	6.86	1.18	78.1	15.15	HiSeq4000	-		DRR208422
MP2_017	MP2_017	3.89	2.96	1.48	0.36	67.0	3.83	HiSeq4000	-		DRR208423
MP2_018	MP2_018	12.70	11.17	7.04	1.10	78.3	15.53	HiSeq4000	-		DRR208424
MP2_019	MP2_019	5.00	4.31	2.32	0.41	74.2	5.38	HiSeq4000	-		DRR208425
MP2_020	MP2_020	10.13	9.04	6.04	0.78	78.1	13.34	HiSeq4000	-		DRR208426
MP2_023	MP2_023	4.98	3.90	2.10	0.35	71.4	5.08	HiSeq4000	-		DRR208427
MP2_024	MP2_024	10.08	8.74	5.10	1.27	75.4	11.68	HiSeq4000	-		DRR208428
MP2_025	MP2_025	4.80	3.53	1.91	0.38	70.2	4.70	HiSeq4000	-		DRR208429
MP2_026	MP2_026	8.36	7.38	4.88	0.66	77.5	10.86	HiSeq4000	-		DRR208430
MP2_027	MP2_027	5.35	3.86	2.05	0.37	71.6	4.93	HiSeq4000	-		DRR208431
MP2_028	MP2_028	8.11	7.08	4.45	0.72	76.4	10.05	HiSeq4000	-		DRR208432
MP2_029	MP2_029	9.89	8.61	5.03	1.08	75.4	11.52	HiSeq4000	-		DRR208433
MP2_031	MP2_031	10.33	9.08	6.04	0.79	78.5	13.30	HiSeq4000	-		DRR208434
MP2_032	MP2_032	16.56	12.57	6.45	1.21	78.9	14.12	HiSeq4000	-		DRR208435
MP2_033	MP2_033	7.32	6.41	4.19	0.62	77.5	9.34	HiSeq4000	-		DRR208436
MP2_034	MP2_034	8.05	6.99	4.40	0.79	75.0	10.12	HiSeq4000	-		DRR208437
MP2_035	MP2_035	9.06	7.95	4.96	0.83	77.3	11.07	HiSeq4000	-		DRR208438
MP2_037	MP2_037	9.70	8.41	5.16	0.99	77.3	11.53	HiSeq4000	-		DRR208439
MP2_039	MP2_039	7.54	6.58	4.00	0.82	75.4	9.17	HiSeq4000	-		DRR208440
MP2_043	MP2_043	9.15	7.93	4.24	0.71	77.3	9.46	HiSeq4000	-		DRR208441
MP2_044	MP2_044	9.75	8.60	5.28	0.95	76.9	11.85	HiSeq4000	-		DRR208442
MP2_047	MP2_047	8.95	7.64	4.04	0.76	77.1	9.03	HiSeq4000	-		DRR208443
MP2_048	MP2_048	8.27	7.24	3.94	0.69	77.4	8.80	HiSeq4000	-		DRR208444
MP2_050	MP2_050	11.17	9.77	5.67	1.35	76.2	12.85	HiSeq4000	-		DRR208445
MP2_052	MP2_052	9.98	8.75	5.18	1.13	75.1	11.90	HiSeq4000	-		DRR208446
MP2_053	MP2_053	11.85	9.88	4.74	2.21	72.0	11.37	HiSeq4000	-		DRR208447
MP2_054	MP2_054	10.38	6.95	3.67	0.70	77.1	8.21	HiSeq4000	-		DRR208448
MP2_055	MP2_055	12.74	10.66	5.55	1.85	74.8	12.81	HiSeq4000	-		DRR208449

MP2_057	MP2_057	8.68	7.41	4.06	1.24	72.2	9.72	HiSeq4000	-	DRR208450
MP2_058	MP2_058	11.14	9.54	6.10	0.89	78.2	13.47	HiSeq4000	-	DRR208451
MP2_060	MP2_060	8.31	7.05	3.51	0.79	76.0	7.97	HiSeq4000	-	DRR208452
MP2_061	MP2_061	12.07	10.38	6.88	0.95	79.0	15.04	HiSeq4000	-	DRR208453
MP2_063	MP2_063	7.03	5.43	2.96	0.51	76.3	6.71	HiSeq4000	-	DRR208454
MP2_064	MP2_064	11.23	9.50	5.46	1.28	76.0	12.39	HiSeq4000	-	DRR208455
MP2_113	MP2_113	6.79	5.71	3.29	0.79	75.0	7.57	HiSeq4000	-	DRR208456
MP2_114	MP2_114	7.80	6.62	3.60	0.94	70.9	8.75	HiSeq4000	-	DRR208457
MP2_116	MP2_116	7.17	6.14	3.78	0.66	75.5	8.64	HiSeq4000	-	DRR208458
MP2_117	MP2_117	6.52	5.53	3.38	0.55	75.9	7.69	HiSeq4000	-	DRR208459
MP2_121	MP2_121	11.64	10.04	5.72	1.45	76.1	12.96	HiSeq4000	-	DRR208460
MP2_122	MP2_122	9.07	7.65	4.33	1.15	75.5	9.89	HiSeq4000	-	DRR208461
MP2_125	MP2_125	9.25	8.04	4.87	0.86	77.7	10.82	HiSeq4000	-	DRR208462
MP2_126	MP2_126	8.65	7.46	4.36	1.00	76.1	9.89	HiSeq4000	-	DRR208463
MP2_127	MP2_127	11.45	9.94	6.22	0.99	78.0	13.76	HiSeq4000	-	DRR208464
MP2_128	MP2_128	10.17	8.91	5.41	1.01	77.1	12.11	HiSeq4000	-	DRR208465
MP2_129	MP2_129	11.75	10.05	5.97	1.32	77.4	13.30	HiSeq4000	-	DRR208466
MP2_130	MP2_130	9.04	7.78	4.94	0.75	76.8	11.10	HiSeq4000	-	DRR208467
MP2_131	MP2_131	10.02	8.69	5.59	0.85	78.2	12.34	HiSeq4000	-	DRR208468
MP2_132	MP2_132	9.93	8.56	5.23	0.99	77.2	11.69	HiSeq4000	-	DRR208469
MP2_133	MP2_133	7.97	6.87	4.29	0.71	77.0	9.63	HiSeq4000	-	DRR208470
MP2_136	MP2_136	9.56	8.20	4.48	1.48	76.2	10.14	HiSeq4000	-	DRR208471
MP2_137	MP2_137	10.99	9.51	5.70	1.15	76.5	12.86	HiSeq4000	-	DRR208472
MP2_138	MP2_138	8.51	7.42	4.61	0.76	77.3	10.28	HiSeq4000	-	DRR208473
MP2_139	MP2_139	9.41	8.27	5.12	0.83	75.9	11.65	HiSeq4000	-	DRR208474
MP2_140	MP2_140	8.91	7.74	4.74	0.90	76.9	10.65	HiSeq4000	-	DRR208475
MP2_141	MP2_141	9.22	7.61	4.05	1.22	72.2	9.69	HiSeq4000	-	DRR208476
MP2_142	MP2_142	10.72	9.12	4.11	2.49	73.3	9.67	HiSeq4000	-	DRR208477
MP2_143	MP2_143	7.99	6.94	4.03	0.91	75.3	9.24	HiSeq4000	-	DRR208478
MP2_144	MP2_144	9.30	8.14	5.31	0.79	77.5	11.83	HiSeq4000	-	DRR208479
MP2_145	MP2_145	10.35	8.99	5.13	1.17	76.5	11.56	HiSeq4000	-	DRR208480
MP2_146	MP2_146	10.87	9.44	5.39	1.41	77.1	12.07	HiSeq4000	-	DRR208481
MP2_147	MP2_147	9.96	8.80	5.79	0.76	78.4	12.75	HiSeq4000	-	DRR208482
MP2_149	MP2_149	9.80	8.64	5.74	0.78	78.0	12.71	HiSeq4000	-	DRR208483
MP2_150	MP2_150	7.47	6.31	3.17	1.22	71.5	7.65	HiSeq4000	-	DRR208484
MP2_151	MP2_151	8.96	7.85	4.80	0.90	78.0	10.63	HiSeq4000	-	DRR208485
MP2_152	MP2_152	12.30	10.66	6.41	1.29	78.8	14.02	HiSeq4000	-	DRR208486
MP2_154	MP2_154	9.78	8.41	4.56	1.42	75.8	10.38	HiSeq4000	-	DRR208487
MP2_155	MP2_155	10.40	9.01	5.31	1.23	77.5	11.82	HiSeq4000	-	DRR208488
MP2_156	MP2_156	8.67	7.49	4.32	1.00	76.2	9.79	HiSeq4000	-	DRR208489
MP2_157	MP2_157	7.64	6.64	4.00	0.84	76.0	9.08	HiSeq4000	-	DRR208490
MP2_158	MP2_158	8.84	7.67	4.85	0.79	77.8	10.77	HiSeq4000	-	DRR208491
MP2_159	MP2_159	9.82	8.47	4.97	1.16	77.2	11.10	HiSeq4000	-	DRR208492
MP2_160	MP2_160	8.43	7.33	4.57	0.73	77.2	10.23	HiSeq4000	-	DRR208493
MP2_161	MP2_161	8.93	7.71	4.46	1.10	77.1	9.99	HiSeq4000	-	DRR208494
MP2_162	MP2_162	12.11	10.46	5.71	1.62	77.4	12.73	HiSeq4000	-	DRR208495
MP2_166	MP2_166	12.03	10.49	6.27	1.21	76.7	14.09	HiSeq4000	-	DRR208496
MP2_167	MP2_167	9.67	8.39	4.63	1.31	74.7	10.70	HiSeq4000	-	DRR208497
MP2_168	MP2_168	15.43	13.47	8.68	1.28	79.0	18.96	HiSeq4000	-	DRR208498

MP2_169	MP2_169	12.87	11.15	6.58	1.40	77.7	14.62	HiSeq4000	-	DRR208499
MP2_170	MP2_170	13.20	11.31	6.24	1.83	77.3	13.94	HiSeq4000	-	DRR208500
MP2_172	MP2_172	11.50	9.60	5.68	1.08	75.6	12.97	HiSeq4000	-	DRR208501
MP2_173	MP2_173	10.20	8.86	4.90	1.31	74.9	11.28	HiSeq4000	-	DRR208502
MP2_174	MP2_174	10.70	9.28	5.37	1.26	77.7	11.95	HiSeq4000	-	DRR208503
MP2_175	MP2_175	13.09	11.51	7.00	1.21	77.4	15.60	HiSeq4000	-	DRR208504
MP2_177	MP2_177	6.33	5.38	2.88	1.00	71.7	6.93	HiSeq4000	-	DRR208505
MP2_178	MP2_178	5.89	5.10	3.00	0.66	73.2	7.07	HiSeq4000	-	DRR208506
MP2_179	MP2_179	4.55	3.89	2.47	0.42	73.5	5.79	HiSeq4000	-	DRR208507
MP2_180	MP2_180	7.09	6.10	3.54	0.86	74.8	8.17	HiSeq4000	-	DRR208508
MP2_181	MP2_181	6.41	5.45	3.05	0.91	72.6	7.26	HiSeq4000	-	DRR208509
MP2_182	MP2_182	8.34	7.16	4.72	0.71	78.2	10.42	HiSeq4000	-	DRR208510
MP2_183	MP2_183	8.89	7.74	5.12	0.74	77.0	11.47	HiSeq4000	-	DRR208511
MP2_185	MP2_185	6.46	5.49	3.06	0.97	72.4	7.30	HiSeq4000	-	DRR208512
MP2_186	MP2_186	6.37	5.37	3.39	0.59	76.0	7.70	HiSeq4000	-	DRR208513
MP2_187	MP2_187	5.86	4.97	2.86	0.72	72.4	6.83	HiSeq4000	-	DRR208514
MP2_188	MP2_188	8.36	7.11	4.48	0.83	76.4	10.12	HiSeq4000	-	DRR208515
MP2_189	MP2_189	6.63	5.69	3.34	0.75	73.9	7.80	HiSeq4000	-	DRR208516
MP2_190	MP2_190	6.41	5.35	3.44	0.58	77.4	7.67	HiSeq4000	-	DRR208517
MP2_191	MP2_191	7.46	6.22	3.76	0.85	74.9	8.67	HiSeq4000	-	DRR208518
MP2_192	MP2_192	6.76	5.71	3.54	0.64	74.8	8.16	HiSeq4000	-	DRR208519
MP2_193	MP2_193	9.63	8.56	5.41	0.86	77.5	12.06	HiSeq4000	-	DRR208520
MP2_196	MP2_196	11.11	9.85	6.23	0.96	78.2	13.76	HiSeq4000	-	DRR208521
MP2_197	MP2_197	7.35	6.22	3.96	0.66	76.6	8.92	HiSeq4000	-	DRR208522
MP2_198	MP2_198	8.72	7.48	4.86	0.74	78.2	10.74	HiSeq4000	-	DRR208523
MP2_199	MP2_199	6.66	5.90	3.58	0.69	74.8	8.25	HiSeq4000	-	DRR208524
MP2_200	MP2_200	7.00	6.22	3.99	0.61	75.8	9.08	HiSeq4000	-	DRR208525
MP2_201	MP2_201	8.36	7.17	4.39	0.86	75.4	10.06	HiSeq4000	-	DRR208526
MP2_202	MP2_202	9.03	7.71	3.83	1.87	74.4	8.88	HiSeq4000	-	DRR208527
MP2_203	MP2_203	7.58	6.73	4.06	0.76	76.8	9.12	HiSeq4000	-	DRR208528
MP2_204	MP2_204	10.55	9.21	5.02	1.48	77.2	11.22	HiSeq4000	-	DRR208529
MP2_205	MP2_205	11.71	10.10	6.18	1.22	77.5	13.76	HiSeq4000	-	DRR208530
MP2_206	MP2_206	8.72	7.29	3.94	1.43	74.1	9.16	HiSeq4000	-	DRR208531
MP2_208	MP2_208	11.54	10.28	6.41	1.12	78.2	14.16	HiSeq4000	-	DRR208532
MP2_211	MP2_211	9.81	8.70	5.44	1.02	78.4	11.98	HiSeq4000	-	DRR208533
MP2_213	MP2_213	10.05	8.77	5.30	1.02	78.0	11.73	HiSeq4000	-	DRR208534
MP2_214	MP2_214	8.64	7.69	4.64	0.96	76.1	10.53	HiSeq4000	-	DRR208535
MP2_215	MP2_215	9.92	8.76	5.62	0.81	78.0	12.43	HiSeq4000	-	DRR208536
MP2_216	MP2_216	9.92	8.64	5.19	1.10	75.4	11.88	HiSeq4000	-	DRR208537
MP2_218	MP2_218	9.62	8.52	5.24	1.10	75.4	11.99	HiSeq4000	-	DRR208538
MP2_219	MP2_219	7.57	6.57	4.15	0.70	74.8	9.57	HiSeq4000	-	DRR208539
MP2_220	MP2_220	7.81	6.90	4.21	0.78	76.1	9.55	HiSeq4000	-	DRR208540
MP2_221	MP2_221	9.33	8.28	5.13	0.92	76.2	11.63	HiSeq4000	-	DRR208541
MP2_222	MP2_222	9.13	7.90	4.79	1.02	75.7	10.93	HiSeq4000	-	DRR208542
MP2_224	MP2_224	11.19	9.85	6.23	1.05	77.1	13.95	HiSeq4000	-	DRR208543
MP2_225	MP2_225	8.97	7.74	4.41	1.09	74.2	10.25	HiSeq4000	-	DRR208544
MP2_227	MP2_227	14.19	12.43	7.97	1.15	78.7	17.48	HiSeq4000	-	DRR208545
MP2_228	MP2_228	9.03	7.86	4.92	0.90	76.8	11.05	HiSeq4000	-	DRR208546
MP2_229	MP2_229	10.39	9.13	5.71	0.97	77.5	12.73	HiSeq4000	-	DRR208547

MP2_231	MP2_231	10.31	8.99	5.62	0.96	77.6	12.50	HiSeq4000	-	DRR208548
MP2_232	MP2_232	11.06	9.64	6.00	1.04	77.1	13.41	HiSeq4000	-	DRR208549
MP2_233	MP2_233	9.57	8.46	5.23	1.07	76.8	11.76	HiSeq4000	-	DRR208550
MP2_234	MP2_234	6.96	6.02	3.42	0.89	73.4	8.05	HiSeq4000	-	DRR208551
MP2_235	MP2_235	8.71	7.54	4.21	1.25	73.9	9.82	HiSeq4000	-	DRR208552
MP2_236	MP2_236	5.82	4.95	3.06	0.56	73.8	7.16	HiSeq4000	-	DRR208553
MP2_237	MP2_237	6.46	5.55	3.27	0.80	74.2	7.61	HiSeq4000	-	DRR208554
MP2_239	MP2_239	7.08	6.14	3.77	0.73	75.0	8.66	HiSeq4000	-	DRR208555
MP2_240	MP2_240	6.92	6.00	3.70	0.78	74.4	8.59	HiSeq4000	-	DRR208556
MP2_241	MP2_241	10.28	8.87	4.73	1.60	74.7	10.92	HiSeq4000	-	DRR208557
MP2_242	MP2_242	8.82	7.65	4.62	0.85	75.3	10.58	HiSeq4000	-	DRR208558
MP2_245	MP2_245	5.90	5.15	3.32	0.51	76.3	7.50	HiSeq4000	-	DRR208559
MP2_246	MP2_246	6.86	5.98	3.77	0.70	76.6	8.50	HiSeq4000	-	DRR208560
MP2_247	MP2_247	6.97	6.01	3.70	0.65	74.3	8.61	HiSeq4000	-	DRR208561
MP2_248	MP2_248	6.45	5.60	3.62	0.57	76.7	8.14	HiSeq4000	-	DRR208562

Table S5. All sequence information of outgroups.

Sample		Fastq size		Aligned bam information				Comment	Accession No.
Name	Name in Scarcelli et al. 2019	Original (Gbp)	Filtered (Gbp)	Aligned (Gbp)	Unmapped (Gbp)	Coverage (%)	Depth		
alata1		28.11	23.95	10.73	1.24	48.0	38.59	D.alata	ERR1019033
alata2		11.58	11.15	3.88	1.37	43.1	15.54	D.alata	SRR7062294
ns004_A5689	A5689	4.22	4.19	3.09	0.34	75.2	7.09	D.abyssinica:Nigeria	SRR8451439
ns005_A5690	A5690	5.79	5.72	4.06	0.37	68.5	10.24	D.abyssinica:Nigeria	SRR8451438
ns006_A5691	A5691	5.53	5.49	2.85	1.73	68.4	7.20	D.abyssinica:Nigeria	SRR8451437
ns007_A5693	A5693	5.93	5.89	4.54	0.15	78.3	10.01	D.abyssinica:Nigeria	SRR8451434
ns008_A5694	A5694	4.87	4.84	3.91	0.04	77.3	8.72	D.abyssinica:Nigeria	SRR8451371
ns009_A5695	A5695	4.55	4.52	3.35	0.42	78.4	7.37	D.abyssinica:Nigeria	SRR8451459
ns010_A5696	A5696	4.75	4.61	3.55	0.22	74.9	8.17	D.abyssinica:Nigeria	SRR8451458
ns011_A5697	A5697	5.70	5.66	4.41	0.15	80.2	9.48	D.abyssinica:Nigeria	SRR8451382
ns012_A5699	A5699	3.25	3.22	2.45	0.15	71.8	5.89	D.abyssinica:Nigeria	SRR8451381
ns013_A5700	A5700	4.79	4.76	3.59	0.32	77.0	8.05	D.abyssinica:Nigeria	SRR8451384
ns014_A5701	A5701	5.99	5.95	4.38	0.37	78.6	9.62	D.abyssinica:Nigeria	SRR8451383
ns015_A5702	A5702	3.96	3.93	2.95	0.29	74.9	6.79	D.abyssinica:Nigeria	SRR8451378
ns016_A5703	A5703	4.53	4.49	3.09	0.37	65.3	8.17	D.abyssinica:Nigeria	SRR8451377
ns017_A5704	A5704	4.95	4.91	2.85	1.17	69.6	7.08	D.abyssinica:Nigeria	SRR8451380
ns018_A5705	A5705	5.54	5.49	3.75	0.67	74.5	8.68	D.abyssinica:Nigeria	SRR8451379
ns019_A52	A52	1.66	1.63	1.44	0.02	70.8	3.52	D.abyssinica:Benin	SRR8451376
ns020_A62	A62	2.35	2.31	2.06	0.02	77.3	4.60	D.abyssinica:Benin	SRR8451375
ns021_A67	A67	7.54	7.42	6.12	0.12	85.2	12.40	D.abyssinica:Benin	SRR8451343
ns023_A467	A467	5.72	5.64	5.08	0.06	82.0	10.69	D.abyssinica:Benin	SRR8451345
ns024_A537	A537	6.22	6.13	5.28	0.05	79.3	11.49	D.abyssinica:Benin	SRR8451346
ns025_A3009	A3009	3.33	3.27	2.92	0.03	76.7	6.57	D.abyssinica:Benin	SRR8451347
ns027_A5068	A5068	1.98	1.95	1.67	0.04	65.7	4.38	D.abyssinica:Ghana	SRR8451349
ns028_A5045	A5045	2.61	2.56	2.21	0.04	74.4	5.12	D.abyssinica:Ghana	SRR8451350
ns029_A5047	A5047	3.32	3.27	2.80	0.04	75.0	6.46	D.abyssinica:Ghana	SRR8451351
ns030_A5048	A5048	9.39	9.23	7.75	0.10	82.9	16.14	D.abyssinica:Ghana	SRR8451352
ns031_A5059	A5059	10.28	10.10	7.09	1.66	82.5	14.82	D.abyssinica:Ghana	SRR8451320
ns032_A5061	A5061	2.81	2.77	1.91	0.54	72.4	4.55	D.abyssinica:Ghana	SRR8451319
ns033_A5066	A5066	8.09	7.95	6.74	0.11	80.7	14.41	D.abyssinica:Ghana	SRR8451318
ns034_A5067	A5067	7.67	7.55	6.51	0.06	82.0	13.71	D.abyssinica:Ghana	SRR8451317
ns035_P5344	P5344	3.33	3.30	2.46	0.10	70.6	6.02	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451316
ns036_P5350	P5350	4.06	4.02	2.77	0.20	63.5	7.52	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451315
ns037_P5358	P5358	4.21	4.17	3.09	0.15	73.2	7.29	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451314
ns038_P5369	P5369	3.10	3.08	2.17	0.32	70.2	5.34	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451313
ns039_P5378	P5378	3.01	2.99	2.31	0.05	70.5	5.66	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451322
ns040_P5381	P5381	3.90	3.87	2.97	0.11	72.8	7.05	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451321

ns041_P5404	P5404	4.53	4.49	3.31	0.31	74.3	7.69	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451462
ns042_P5413	P5413	3.78	3.75	2.82	0.16	73.5	6.62	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451463
ns043_P5417	P5417	4.61	4.58	3.44	0.19	74.1	8.01	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451460
ns044_P5420	P5420	2.25	2.23	1.65	0.15	65.9	4.31	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451461
ns045_P5424	P5424	5.30	5.26	3.74	0.42	74.4	8.68	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451466
ns046_P5427	P5427	4.25	4.22	3.24	0.05	72.9	7.66	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451467
ns047_P5430	P5430	3.34	3.31	2.41	0.10	63.5	6.56	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451464
ns048_P5434	P5434	2.80	2.77	2.10	0.06	61.8	5.86	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451465
ns049_P5438	P5438	3.64	3.61	2.36	0.62	70.6	5.76	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451468
ns050_P5441	P5441	4.13	4.09	3.04	0.23	73.7	7.12	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451469
ns051_P5448	P5448	4.73	4.69	3.66	0.09	73.6	8.58	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451449
ns054_P5318	P5318	5.04	4.99	3.07	0.62	67.7	7.83	D.praehensilis:Cameroon:Cameroonian D.praehensilis	SRR8451450
ns055_P5746	P5746	3.80	3.77	2.66	0.43	65.3	7.02	D.praehensilis:Nigeria:Western D.praehensilis	SRR8451453
ns056_P5708	P5708	6.19	6.13	4.22	0.39	64.5	11.30	D.praehensilis:Nigeria:Western D.praehensilis	SRR8451452
ns057_P5710	P5710	3.89	3.86	2.61	0.48	70.0	6.42	D.praehensilis:Nigeria:Western D.praehensilis	SRR8451455
ns058_P5713	P5713	3.24	3.21	2.34	0.22	67.2	6.02	D.praehensilis:Nigeria:Western D.praehensilis	SRR8451454
ns059_P5716	P5716	2.56	2.53	1.91	0.03	63.0	5.23	D.praehensilis:Nigeria:Western D.praehensilis	SRR8451457
ns061_P5720	P5720	3.87	3.84	2.99	0.17	73.5	7.02	D.praehensilis:Nigeria:Western D.praehensilis	SRR8451430
ns062_P5723	P5723	3.63	3.61	2.17	0.93	68.9	5.44	D.praehensilis:Nigeria:Western D.praehensilis	SRR8451431
ns063_P5728	P5728	3.75	3.71	2.65	0.34	64.3	7.11	D.praehensilis:Nigeria:Western D.praehensilis	SRR8451432
ns064_P5729	P5729	7.31	7.25	4.58	1.01	72.5	10.89	D.praehensilis:Nigeria:Western D.praehensilis	SRR8451433
ns065_P424	P424	3.46	3.40	3.03	0.04	79.1	6.61	D.praehensilis:Benin:Western D.praehensilis	SRR8451426
ns066_P425	P425	1.63	1.60	1.44	0.02	69.5	3.57	D.praehensilis:Benin:Western D.praehensilis	SRR8451427
ns067_P457	P457	4.21	4.13	3.46	0.12	74.5	8.01	D.praehensilis:Benin:Western D.praehensilis	SRR8451428
ns068_P462	P462	4.33	4.26	3.68	0.08	79.7	7.98	D.praehensilis:Benin:Western D.praehensilis	SRR8451429
ns069_P323	P323	4.22	4.15	3.70	0.05	80.5	7.94	D.praehensilis:Benin:Western D.praehensilis	SRR8451435
ns070_P464	P464	5.29	5.21	4.65	0.05	80.6	9.96	D.praehensilis:Benin:Western D.praehensilis	SRR8451436
ns073_P2990	P2990	2.88	2.84	2.56	0.03	77.6	5.70	D.praehensilis:Benin:Western D.praehensilis	SRR8451409
ns075_P4918	P4918	2.45	2.40	1.82	0.27	72.6	4.33	D.praehensilis:Ghana:Western D.praehensilis	SRR8451415
ns076_P4919	P4919	5.46	5.36	4.04	0.45	79.4	8.79	D.praehensilis:Ghana:Western D.praehensilis	SRR8451414
ns077_P4920	P4920	6.04	5.93	4.63	0.53	80.3	9.95	D.praehensilis:Ghana:Western D.praehensilis	SRR8451413
ns078_P4921	P4921	4.73	4.65	3.73	0.31	79.5	8.11	D.praehensilis:Ghana:Western D.praehensilis	SRR8451412
ns079_P4928	P4928	3.77	3.71	2.99	0.24	78.4	6.57	D.praehensilis:Ghana :Western D.praehensilis	SRR8451407

ACKNOWLEDGEMENTS

The author thanks the collaborators, Kwabena Darkwa, Hiroki Yaegashi, Satoshi Natsume, Motoki Shimizu, Akira Abe, Akiko Hirabuchi, Kazue Ito, Kaori Oikawa, Muluneh Tamiru-Oli, Atsushi Ohta, Ryo Matsumoto, Agre Paterne, David De Koeyer, Babil Pachakkil, Shinsuke Yamanaka, Satoru Muranaka, Hiroko Takagi, Ben White, Robert Asiedu, Hideki Innan, Asrat Asfaw, Patrick Adebola, Aoi Kudoh, Ryohei Terauchi for carrying out the research.

This study was carried out mainly at Iwate Biotechnology Research Center (IBRC) under the AfricaYam Project funded by the Bill and Melinda Gates Foundation (BMGF) as well as the EDITS-Yam project funded by JIRCAS, Japan. The author also thanks Yoshitaka Takano, Kentaro Yoshida and Sophien Kamoun for valuable comments on the thesis.

REREFENCES

Akakpo, R., Scarcelli, N., Chaïr, H., Dansi, A., Djedatin, G., Thuillet, A.-C., Rhoné, B., François, O., Alix, K., & Vigouroux, Y. (2017). Molecular basis of African yam domestication: Analyses of selection point to root development, starch biosynthesis, and photosynthesis related genes. *BMC Genomics*, *18*(1), 782. https://doi.org/10.1186/s12864-017-4143-2

Arnau, G., Abraham, K., Sheela, M. N., Chair, H., Sartie, A., & Asiedu, R. (2010). Yams. In *Root and Tuber Crops* (pp. 127–148). Springer, New York, NY. https://link.springer.com/chapter/10.1007/978-0-387-92765-7_4

Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Broman, K. W., Wu, H., Sen, Ś., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, *19*(7), 889–890. https://doi.org/10.1093/bioinformatics/btg112

Browning, S. R., & Browning, B. L. (2007). Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *The American Journal of Human Genetics*, *81*(5), 1084–1097. https://doi.org/10.1086/521987

Burkill, I. H. (1960). The organography and the evolution of Dioscoreaceae, the family of the yams. *Journal of the Linnean Society*, *56*, 319–412.

Cabanettes, F., & Klopp, C. (2018). D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. *PeerJ*, *6*, e4958. https://doi.org/10.7717/peerj.4958

Caddick, L. R., Wilkin, P., Rudall, P. J., Hedderson, T. A. J., & Chase, M. W. (2002). Yams reclassified: A recircumscription of Dioscoreaceae and Dioscoreales. *TAXON*, *51*(1), 103–114. https://doi.org/10.2307/1554967

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Chaïr, H., Cornet, D., Deu, M., Baco, M. N., Agbangla, A., Duval, M. F., & Noyer, J. L. (2010). Impact of farmer selection on yam genetic diversity. *Conservation Genetics*, *11*(6), 2255–2265. https://doi.org/10.1007/s10592-010-0110-z

Chaïr, H., Sardos, J., Supply, A., Mournet, P., Malapa, R., & Lebot, V. (2016). Plastid phylogenetics of Oceania yams (Dioscorea spp., Dioscoreaceae) reveals natural interspecific hybridization of the greater yam (D. alata). *Botanical Journal of the Linnean Society*, *180*(3), 319–333. https://doi.org/10.1111/boj.12374

Cormier, F., Lawac, F., Maledon, E., Gravillon, M.-C., Nudol, E., Mournet, P., Vignes, H., Chaïr, H., & Arnau, G. (2019). A reference high-density genetic map of greater yam (Dioscorea alata L.). *Theoretical and Applied Genetics*, *132*(6), 1733–1744. https://doi.org/10.1007/s00122-019-03311-6

Coursey, D. G. (1967). Yams. An account of the nature, origins, cultivation and utilisation of the useful members of the Dioscoreaceae. In *Tropical agricultural series* (pp. 108–129). London: Longmans, Green and Co. Ltd.

Coursey, D. G. (1972). The Civilizations of the Yam: Interrelationships of Man and Yams in Africa and the Indo-Pacific Region. *Archaeology & Physical Anthropology in Oceania*, *7*(3), 215–233.

Coursey, D. G. (1976a). The origins and domestication of yams in Africa. In *Origins of African Plant Domestication* (J. R. Harlan, J. M. J. D. Wet, A. B. L. Stemler, Eds, pp. 383–408). De Gruyter Mouton.

Coursey, D. G. (1976b). Yams: Dioscorea spp. (Dioscoreaceae). In *Evolution of Crop Plants* (N. W. Simmonds, ed, pp. 70–74). Longman Group.

Couto, R. S., Martins, A. C., Bolson, M., Lopes, R. C., Smidt, E. C., & Braga, J. M. A. (2018). Time calibrated tree of Dioscorea (Dioscoreaceae) indicates four origins of yams in the Neotropics since the Eocene. *Botanical Journal of the Linnean Society*, *188*(2), 144–160. https://doi.org/10.1093/botlinnean/boy052

Darkwa, K., Olasanmi, B., Asiedu, R., & Asfaw, A. (2020). Review of empirical and emerging breeding methods and tools for yam (Dioscorea spp.) improvement: Status and prospects. *Plant Breeding*, *139*(3), 474–497. https://doi.org/10.1111/pbr.12783

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics*, *34*(15), 2666–2669. https://doi.org/10.1093/bioinformatics/bty149

Dutta, B. (2015). *Food and medicinal values of certain species of Dioscorea with special reference to Assam*. 5.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust Demographic Inference from Genomic and SNP Data. *PLOS Genetics*, *9*(10), e1003905. https://doi.org/10.1371/journal.pgen.1003905

FAOSTAT. (2018). Food and Agriculture Organization of the United Nations. http://www.fao.org/statistics/

Folk, R. A., Soltis, P. S., Soltis, D. E., & Guralnick, R. (2018). New prospects in the detection and comparative analysis of hybridization in the tree of life. *American Journal of Botany*, *105*(3), 364–375. https://doi.org/10.1002/ajb2.1018

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and Efficient Estimation of Individual Ancestry Coefficients. *Genetics*, *196*(4), 973–983. https://doi.org/10.1534/genetics.113.160572

Girma, G., Hyma, K. E., Asiedu, R., Mitchell, S. E., Gedil, M., & Spillane, C. (2014). Next-generation sequencing based genotyping, cytometry and phenotyping for understanding diversity and evolution of guinea yams. *Theoretical and Applied Genetics*, *127*(8), 1783–1794. https://doi.org/10.1007/s00122-014-2339-2

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLOS Genetics*, *5*(10), e1000695. https://doi.org/10.1371/journal.pgen.1000695

Hancock, J. F. (2012). *Plant evolution and the origin of crop species* (ed. 3). CABI Publishing.

Heslop-Harrison, J. S., & Schwarzacher, T. (2007). Domestication, Genomics and the Future for Banana. *Annals of Botany*, *100*(5), 1073–1084. https://doi.org/10.1093/aob/mcm191

Hu, X.-S., & Filatov, D. A. (2016). The large-X effect in plants: Increased species divergence and reduced gene flow on the Silene X-chromosome. *Molecular Ecology*, *25*(11), 2609–2619. https://doi.org/10.1111/mec.13427

Hughes, C. E., Govindarajulu, R., Robertson, A., Filer, D. L., Harris, S. A., & Bailey, C. D. (2007). Serendipitous backyard hybridization and the origin of crops. *Proceedings of the National Academy of Sciences*, *104*(36), 14389–14394. https://doi.org/10.1073/pnas.0702193104

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

Ikiriza, H., Ogwang, P. E., Peter, E. L., Hedmon, O., Tolo, C. U., Abubaker, M., & Abdalla, A. A. M. (2019). Dioscorea bulbifera, a highly threatened African medicinal plant, a review. *Cogent Biology*, *5*(1), 1631561. https://doi.org/10.1080/23312025.2019.1631561

Iwata, H., & Gotoh, O. (2012). Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. *Nucleic Acids Research*, *40*(20), e161. https://doi.org/10.1093/nar/gks708

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), Article 4. https://doi.org/10.1038/nmeth.3317

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, *37*(5), Article 5. https://doi.org/10.1038/s41587-019-0072-8

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Leigh, J. W., & Bryant, D. (2015). popart: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liu, X.-T., Wang, Z.-Z., Xiao, W., Zhao, H.-W., Hu, J., & Yu, B. (2008). Cholestane and spirostane glycosides from the rhizomes of Dioscorea septemloba. *Phytochemistry*, *69*(6), 1411–1418. https://doi.org/10.1016/j.phytochem.2007.12.014

Lo, C.-C., & Chain, P. S. G. (2014). Rapid evaluation and quality control of next generation sequencing data with FaQCs. *BMC Bioinformatics*, *15*(1), 366. https://doi.org/10.1186/s12859-014-0366-2

Magwé-Tindo, J., Wieringa, J. J., Sonké, B., Zapfack, L., Vigouroux, Y., Couvreur, T. L. P., & Scarcelli, N. (2018). Guinea yam (Dioscorea spp., Dioscoreaceae) wild relatives identified using whole plastome phylogenetic analyses. *TAXON*, *67*(5), 905–915. https://doi.org/10.12705/675.4

Mallet, J. (2007). Hybrid speciation. *Nature*, *446*(7133), Article 7133. https://doi.org/10.1038/nature05706

Maurin, O., Muasya, A. M., Catalan, P., Shongwe, E. Z., Viruel, J., Wilkin, P., & van der Bank, M. (2016). Diversification into novel habitats in the Africa clade of Dioscorea (Dioscoreaceae): Erect habit and elephant's foot tubers. *BMC Evolutionary Biology*, *16*(1), 238. https://doi.org/10.1186/s12862-016-0812-z

McCauley, D. E. (1995). The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends in Ecology & Evolution*, *10*(5), 198–202. https://doi.org/10.1016/S0169-5347(00)89052-7

Murty, Y. S. & Purnima. (1983). Morphology, anatomy and development of bulbil in some dioscoreas. *Proceedings / Indian Academy of Sciences*, *92*(6), 443–449. https://doi.org/10.1007/BF03053017

Nei, M., & Tajima, F. (1981). DNA POLYMORPHISM DETECTABLE BY RESTRICTION ENDONUCLEASES. *Genetics*, *97*(1), 145–163. https://doi.org/10.1093/genetics/97.1.145

Niknafs, Y. S., Pandian, B., Iyer, H. K., Chinnaiyan, A. M., & Iyer, M. K. (2017). TACO produces robust multisample transcriptome assemblies from RNA-seq. *Nature Methods*, *14*(1), Article 1. https://doi.org/10.1038/nmeth.4078

Noda, H., Yamashita, J., Fuse, S., Pooma, R., Poopath, M., Tobe, H., & Tamura, M. N. (2020). A Large-scale Phylogenetic Analysis of *Dioscorea* (Dioscoreaceae), with Reference to Character Evolution and Subgeneric Recognition. *Acta Phytotaxonomica et Geobotanica*, *71*(2), 103–128. https://doi.org/10.18942/apg.201923

Obidiegwu, J. E., & Akpabio, E. M. (2017). The geography of yam cultivation in southern Nigeria: Exploring its social meanings and cultural functions. *Journal of Ethnic Foods*, *4*(1), 28–35. https://doi.org/10.1016/j.jef.2017.02.004

Obidiegwu, J. E., Lyons, J. B., & Chilaka, C. A. (2020). The Dioscorea Genus (Yam)—An Appraisal of Nutritional and Therapeutic Potentials. *Foods*, *9*(9), Article 9. https://doi.org/10.3390/foods9091304

Peng, J. H., Sun, D., & Nevo, E. (2011). Domestication evolution, genetics and genomics in wheat. *Molecular Breeding*, *28*(3), 281–301. https://doi.org/10.1007/s11032-011-9608-4

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), Article 3. https://doi.org/10.1038/nbt.3122

Peter, B. M. (2016). Admixture, Population Structure, and F-Statistics. *Genetics*, *202*(4), 1485–1501. https://doi.org/10.1534/genetics.115.183913

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

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), Article 6. https://doi.org/10.1038/ng.3845

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, *461*(7263), Article 7263. https://doi.org/10.1038/nature08365

Rieseberg, L. H. (1991). Homoploid Reticulate Evolution in Helianthus (asteraceae): Evidence from Ribosomal Genes. *American Journal of Botany*, *78*(9), 1218–1237. https://doi.org/10.1002/j.1537-2197.1991.tb11415.x

Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, *19*(1), 460. https://doi.org/10.1186/s12859-018-2485-7

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

Salman-Minkov, A., Sabath, N., & Mayrose, I. (2016). Whole-genome duplication as a key factor in crop domestication. *Nature Plants*, *2*(8), Article 8. https://doi.org/10.1038/nplants.2016.115

Scarcelli, N., Chaïr, H., Causse, S., Vesta, R., Couvreur, T. L. P., & Vigouroux, Y. (2017). Crop wild relative conservation: Wild yams are not that wild. *Biological Conservation*, *210*, 325–333. https://doi.org/10.1016/j.biocon.2017.05.001

Scarcelli, N., Cubry, P., Akakpo, R., Thuillet, A.-C., Obidiegwu, J., Baco, M. N., Otoo, E., Sonké, B., Dansi, A., Djedatin, G., Mariac, C., Couderc, M., Causse, S., Alix, K., Chaïr, H., François, O., & Vigouroux, Y. (2019). Yam genomics supports West Africa as a major cradle of crop domestication. *Science Advances*, *5*(5), eaaw1947. https://doi.org/10.1126/sciadv.aaw1947

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

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, *27*(6), 863–864. https://doi.org/10.1093/bioinformatics/btr026

Sharif, B. M., Burgarella, C., Cormier, F., Mournet, P., Causse, S., Van, K. N., Kaoh, J., Rajaonah, M. T., Lakshan, S. R., Waki, J., Bhattacharjee, R., Badara, G., Pachakkil, B., Arnau, G., & Chaïr, H. (2020). Genome-wide genotyping elucidates the geographical diversification and dispersal of the polyploid and clonally propagated yam (Dioscorea alata). *Annals of Botany*, *126*(6), 1029–1038. https://doi.org/10.1093/aob/mcaa122

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Sugihara, Y., Darkwa, K., Yaegashi, H., Natsume, S., Shimizu, M., Abe, A., Hirabuchi, A., Ito, K., Oikawa, K., Tamiru-Oli, M., Ohta, A., Matsumoto, R., Agre, P., Koeyer, D. D., Pachakkil, B., Yamanaka, S., Muranaka, S., Takagi, H., White, B., … Terauchi, R. (2020). Genome analyses reveal the hybrid origin of the staple crop white Guinea yam (Dioscorea rotundata). *Proceedings of the National Academy of Sciences*, *117*(50), 31987–31992. https://doi.org/10.1073/pnas.2015830117

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595. https://doi.org/10.1093/genetics/123.3.585

Tamiru, M., Becker, H. C., & Maass, B. L. (2007). Genetic Diversity in Yam Germplasm from Ethiopia and Their Relatedness to the Main Cultivated Dioscorea Species Assessed by AFLP Markers. *Crop Science*, *47*(4), 1744–1753. https://doi.org/10.2135/cropsci2006.11.0719

Tamiru, M., Natsume, S., Takagi, H., White, B., Yaegashi, H., Shimizu, M., Yoshida, K., Uemura, A., Oikawa, K., Abe, A., Urasaki, N., Matsumura, H., Babil, P., Yamanaka, S., Matsumoto, R., Muranaka, S., Girma, G., Lopez-Montes, A., Gedil, M., … Terauchi, R. (2017). Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. *BMC Biology*, *15*(1), 86. https://doi.org/10.1186/s12915-017-0419-x

Terauchi, R., Chikaleke, V. A., Thottappilly, G., & Hahn, S. K. (1992). Origin and phylogeny of Guinea yams as revealed by RFLP analysis of chloroplast DNA and nuclear ribosomal DNA. *Theoretical and Applied Genetics*, *83*(6), 743–751. https://doi.org/10.1007/BF00226693

Terauchi, R., & Kahl, G. (1999). Mapping of the Dioscorea tokoro genome: AFLP markers linked to sex. *Genome*, *42*(4), 752–762. https://doi.org/10.1139/g99-001

Terauchi, R., Terachi, T., & Tsunewaki, K. (1991). Intraspecific variation of chloroplast DNA in Dioscorea bulbifera L. *Theoretical and Applied Genetics*, *81*(4), 461–470. https://doi.org/10.1007/BF00219435

Veyres, N., Aono, M., Sangwan-Norree, B. S., & Sangwan, R. S. (2008a). Has Arabidopsis SWEETIE protein a role in sugar flux and utilization? *Plant Signaling & Behavior*, *3*(9), 722–725. https://doi.org/10.4161/psb.3.9.6470

Veyres, N., Danon, A., Aono, M., Galliot, S., Karibasappa, Y. B., Diet, A., Grandmottet, F., Tamaoki, M., Lesur, D., Pilard, S., Boitel-Conti, M., Sangwan-Norreel, B. S., & Sangwan, R. S. (2008b). The Arabidopsis sweetie mutant is affected in carbohydrate metabolism and defective in the control of

growth, development and senescence. *The Plant Journal*, *55*(4), 665–686. https://doi.org/10.1111/j.1365-313X.2008.03541.x

Viruel, J., Segarra-Moragues, J. G., Raz, L., Forest, F., Wilkin, P., Sanmartín, I., & Catalán, P. (2016). Late Cretaceous–Early Eocene origin of yams (Dioscorea, Dioscoreaceae) in the Laurasian Palaearctic and their subsequent Oligocene–Miocene diversification. *Journal of Biogeography*, *43*(4), 750–762. https://doi.org/10.1111/jbi.12678

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE*, *9*(11), e112963. https://doi.org/10.1371/journal.pone.0112963

Warschefsky, E., Penmetsa, R. V., Cook, D. R., & von Wettberg, E. J. B. (2014). Back to the wilds: Tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. *American Journal of Botany*, *101*(10), 1791–1800. https://doi.org/10.3732/ajb.1400116

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, *7*(2), 256–276. https://doi.org/10.1016/0040-5809(75)90020-9

WCSP. (2020). World Checklist of Selected Plant Families. http://wcsp.science.kew.org

Wilkin, P., Schols, P., Chase, M. W., Chayamarit, K., Furness, C. A., Huysmans, S., Rakotonasolo, F., Smets, E., & Thapyai, C. (2005). A Plastid Gene Phylogeny Of the Yam Genus, Dioscorea: Roots, Fruits and Madagascar. *Systematic Botany*, *30*(4), 736–749. https://doi.org/10.1600/036364405775097879

Wright, S. (1951). THE GENETICAL STRUCTURE OF POPULATIONS. *Annals of Eugenics*, *15*(1), 323–354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x

Wu, Y., Bhat, P. R., Close, T. J., & Lonardi, S. (2008). Efficient and Accurate Construction of Genetic Linkage Maps from the Minimum Spanning Tree of a Graph. *PLOS Genetics*, *4*(10), e1000212. https://doi.org/10.1371/journal.pgen.1000212

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

Zhang, B.-W., Xu, L.-L., Li, N., Yan, P.-C., Jiang, X.-H., Woeste, K. E., Lin, K., Renner, S. S., Zhang, D.-Y., & Bai, W.-N. (2019). Phylogenomics Reveals an Ancient Hybrid Origin of the Persian Walnut. *Molecular Biology and Evolution*, *36*(11), 2451–2461. https://doi.org/10.1093/molbev/msz112

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606