Received: 30 June 2020 | Accepted: 1 December 2020

DOI: 10.1002/csc2.20431

Crop Science

ORIGINAL RESEARCH ARTICLE

Plant Genetic Resources

Simple sequence repeat-based mini-core collection for white Guinea yam (*Dioscorea rotundata*) germplasm

Babil Pachakkil[1,2,*] | **Shinsuke Yamanaka**[1,*] | **Gezahegn Girma**[3,4] | **Ryo Matsumoto**[1,3] | **Muluneh Tamiru-Oli**[5,6] | **Ranjana Bhattacharjee**[3] | **Michael Abberton**[3] | **Satoru Muranaka**[1] | **Robert Asiedu**[3] | **Ryohei Terauchi**[5,7] | **Hiroko Takagi**[1]

[1] Japan International Research Center for Agricultural Sciences, 1-1 Ohwashi, Tsukuba, Ibaraki 305-8686, Japan

[2] Tokyo Univ. of Agriculture, 1-1-1 Sakuragaoka, Setagaya, Tokyo 156-8502, Japan

[3] International Institute of Tropical Agriculture, PMB 5320, Ibadan, Oyo State, Nigeria

[4] Purdue Univ., 915 W State St., West Lafayette, IN 47907, USA

[5] Iwate Biotechnology Research Center, 22-174-4 Narita, Kitakami, Iwate 024-0003, Japan

[6] La Trobe Univ., AgriBio Building, 5 Ring Road, Bundoora, Melbourne VIC 3086, Australia

[7] Kyoto Univ., 1 Nakajou, Mozume-cho, Muko, Kyoto 617-0001, Japan

Correspondence
Shinsuke Yamanaka, Japan International Research Center for Agricultural Sciences, 1-1 Ohwashi, Tsukuba, Ibaraki, 305-8686, Japan.
Email: syamanak@affrc.go.jp

*Babil Pachakkil and Shinsuke Yamanaka contributed equally to this work.

Assigned to Associate Editor Manjit Kang.

Abstract

A core subset with a small number of accessions representing the genetic diversity of the base collection plays a vital role in facilitating efficient utilization of plant genetic resources. This is particularly relevant for vegetatively propagated large plant size tuber crops with a long growing period, such as white Guinea yam (*Dioscorea rotundata* Poir.). For the efficient utilization of *D. rotundata* genetic resources, this study was aimed at developing a mini-core collection from a core collection of 447 *D. rotundata* accessions maintained at the International Institute of Tropical Agriculture (IITA). Accordingly, a *D. rotundata* mini-core collection representing 102 accessions was selected using 16 simple sequence repeat (SSR) markers, retaining ~98% of the SSR allelic diversity of the base collection. A similar level of diversity was captured within the mini-core collection and the base collection with respect to 21 morphological traits, ploidy level, and geographic origin. The mini-core collection demonstrated a wide range of variation in agronomic traits such as growth period, number of tubers, average tuber weight, and total yield per plant. This variation was considerable when compared with the variation observed for the same traits among the 10 lines or genotypes conventionally used in the breeding program at IITA, which were included in this study as checks. The selected mini-core accessions could serve as a working collection to broaden the genetic variation for use in practical breeding programs, as well as in future genomic analyses aimed at the genetic improvement of *D. rotundata* in West Africa.

1 | INTRODUCTION

Yam is a common name for multiple crop species belonging to the genus *Dioscorea* that are widely cultivated as staple crop in tropical and subtropical regions. In West Africa, which accounts for 92% of the world's annual yam production (~67 million tons; FAO, 2018), yam plays an important role

Abbreviations: AMOVA, analysis of molecular variance; DI, DNA index; GRC, Genetic Resources Center; IITA, International Institute of Tropical Agriculture; PCR, polymerase chain reaction; PIC, polymorphic information content; SSR, simple sequence repeat; TBE, tris-borate-ethylenediamine tetra acetic acid.

in regional food security and income generation. *Dioscorea rotundata* Poir. is the most cultivated species in this region, representing majority of the total yam production. However, long growth cycle, inconsistency or no flowering, dioecy, non-synchronous flowering of parental genotypes, polyploidy, and a high level of heterozygosity remain major limitations to genetic research and breeding of *D. rotundata* (Asiedu, Ng, Bai, Ekanayake, & Wanyera, 1998; Girma et al., 2019).

Analysis and utilization of genetic resources with a wide range of variation play a vital role in the genetic improvement of crop plants. For *D. rotundata*, considerable efforts have been made worldwide to collect, maintain, and characterize germplasm, mainly from countries in the "yam belt," including Cote d'Ivoire (Hamon & Bakary, 1990), Ghana (Demuyakor, Dukrog, & Chikpah, 2013), Togo (Dansi et al., 2013), Benin (Dansi et al., 1999), Nigeria (Mignouna et al., 2005), and Cameroon (Mignouna, Dansi, & Zok, 2002). The International Institute of Tropical Agriculture (IITA), based in Ibadan, Nigeria, holds the largest collection of yam genetic resources, with over 3,000 accessions, mainly from West African countries (Girma et al., 2018).

To facilitate efficient evaluation and utilization of genetic resources, Frankel (1984) proposed the concept of a core collection, and defined it as a limited set of accessions representing, with minimum repetitiveness, the genetic diversity of a crop species and its wild relatives. Core collections have been developed for several crops based on geographical origin as well as morphological, agronomic, biochemical, and molecular data (Odong, Jansen, van Eeuwijk, & van Hintum, 2013). An attempt was also made to develop the first West African yam core collection in 2007 (Mahalakshmi et al., 2007). Several years later and after acquiring additional yam germplasm, the IITA Genetic Resources Center (GRC) established a revised core collection consisting of 843 accessions representing six *Dioscorea* spp. based on morphological traits (Girma et al., 2018). The revised core subset consisted of 620 *D. rotundata* accessions. However, managing such a large germplasm collection is a challenge, particularly in yam, because of its long growing period, large plant size, and vegetative propagation. In addition, a core collection of this size remains too large for the detailed analyses of phenotypic and genotypic variations, particularly with respect to quantitative traits that are greatly influenced by environmental factors and require multilocation and multiseason evaluations in replicated trials. To overcome these problems, the development of a core subset for *D. rotundata* is necessary.

The availability of a subset of a core collection based on molecular markers allows an efficient evaluation of traits influenced by environmental factors. Such a molecular-based core subset, commonly known as mini-core collection, has been developed for multiple crop species, including rice (*Oryza sativa* L.; Ebana, Kojima, Fukuoka, Nagamine, & Kawase, 2008) and sorghum [*Sorghum bicolor* (L.) Moench; Shehzad, Okuizumi, Kawase, & Okuno, 2009]. Molecular characterizations of the genetic diversity in *D. rotundata* using DNA markers such as random amplification of polymorphic DNA (RAPD), amplified fragment length polymorphism analysis (AFLP), and simple sequence repeat (SSR) have been reported (Dansi et al., 2000; Mignouna et al., 2002, 2005; Sartie, Asiedu, & Franco, 2012; Tamiru, Becker, & Maass, 2007). However, most of these studies have been restricted to a small number of accessions from limited geographic regions. Therefore, the aims of this study were to assess the genetic diversity in *D. rotundata* accessions maintained at the IITA GRC, which represents the largest collection of this species worldwide, using simple sequence repeat (SSR) markers and to develop a mini-core collection that could aid in efficient utilization of genetic resources in white Guinea yam improvement programs.

Core Ideas

- Mini-core collection for *Dioscorea rotundata* with a small number of accessions.
- We selected 102 accessions of mini-core from the 447 *D. rotundata* core collection based on SSR.
- Mini-core retained ~98% of the SSR variation of the original core collection.
- A similar variation of morphological traits and ploidy level was retained in the mini-core.
- The mini-core collection also demonstrated a wide range of performance in agronomic traits.

2 | MATERIALS AND METHODS

2.1 | Plant materials

Of the 620 *D. rotundata* accessions included in the IITA yam core collection reselected by Girma et al. (2018), 447 accessions were used as the base collection for developing a mini-core collection. The accessions were planted at IITA, Ibadan, Nigeria (7°30′8″ N, 3°54′38″ E). Leaf samples were collected from each accession for DNA extraction and further analysis. For preliminary evaluation of agronomic traits, such as number of stems per plant, growth period, number of tubers per plant, total yield per plant, and average tuber weight of the selected mini-core accessions, 10 lines from the IITA Yam Breeding Unit were used as checks (i.e., "reference lines"). The reference lines comprised five IITA breeding lines (TDr 97/00793, TDr 97/00917, TDr 97/00777, TDr 99/02607, and TDr 95/01932) and five local cultivars or landraces (TDr

EHOBIA, TDr EHURU, TDr OMI EFUN, TDr POUNA, and TDr 04-219).

2.2 | Genomic DNA isolation

Leaf samples of 447 accessions from the base collection available in the field during the leaf tissue sampling period were collected from the IITA field gene bank at Ibadan in September 2012. The collected leaves were lyophilized and kept under cool dry conditions. Genomic DNA was extracted using the CTAB (cetyl trimethylammonium bromide) method (Fulton, Chunwongse, & Tanksley, 1995) with minor modifications. Immediately after grinding of the leaf samples and prior to the addition of the extraction buffer, a washing step was carried out using a 0.1 M HEPES (*N*-2-hydroxyetylpiperazine-*N*′-ethanesulphonic acid) buffer (pH 8.0) containing 1% polyvinylpyrrolidone, 0.9% L-ascorbic acid, and 2% 2-mercaptoethanol to remove polysaccharides and phenolic compounds. DNA was quantified using a GeneQuant Pro spectrophotometer (Amersham Biosciences) and diluted to a working concentration of 20–30 ng μl^{-1} for polymerase chain reaction (PCR) amplification.

2.3 | SSR markers, PCR conditions, and electrophoresis

Ninety SSR markers developed by Tamiru et al. (2015) for yam diversity analysis from a microsatellite-enriched genomic library of yellow Guinea yam (*D. cayenensis* Lam.) were prescreened on 16 randomly selected *D. rotundata* core collection accessions to identify a set of polymorphic markers for further genotyping. A PCR amplification was performed in a 5-µl total reaction volume containing 1× Ex Taq Buffer (TaKaRa), 200 µM dNTPs, 1 µM of each of the forward and reverse primer, 0.25 U Takara Ex Taq HS, and ~20 ng template DNA. The PCR was performed using a PTC-200 thermal cycler (Bio-Rad) with initial denaturation at 94 ˚C for 3 min, followed by 35 cycles of denaturation at 94 ˚C for 30 s, primer annealing at 55 ˚C for 1 min, and extension at 72 ˚C for 30 s, with a final extension step at 72 ˚C for 10 min. The PCR products were electrophoresed on an 8% polyacrylamide gel (mono: bis = 29:1) with a constant power supply of 150 V for 90 min. Tris-borate-ethylenediamine tetraacetic acid (TBE, 10×) and 0.5× TBE buffer were used for gel preparation and electrophoresis, respectively. Gels were stained with ethidium bromide solution, and gel images were taken using the Dolphin-View 2 gel documentation analysis system (Kurabo). Analysis of gel images was performed with Dolphin-1D Gel analysis software version 2.7 (Wealtec Corporation).

2.4 | SSR analysis and informativeness

The SSR alleles detected at the expected product size of each SSR marker were scored in base pairs. As genetic diversity parameters, number of alleles (N_a), observed heterozygosity (H_o), expected heterozygosity (H_e), and polymorphic information content (PIC) were determined using GeneAlEx 6.5 software (Peakall & Smouse, 2012). To reveal population structure of the *D. rotundata* base collection, a Bayesian model clustering algorithm was implemented using SSR data to infer genetic structure and to determine the number of clusters using STRUCTURE version 2.3.4 software (Pritchard, Stephens, & Donnelly, 2000) with an admixture model for ancestry and an independent model for allele frequency, without any prior information about the origin of samples. For each value of K (number of inferred ancestral populations) from 2 to 10, analyses were performed 10 times with 100,000 iterations after a burn-in period of 100,000 iterations. The ΔK was used to estimate the appropriate K value according to the criterion of Evanno, Regnaut, and Goudet (2005). In addition, to determine informativeness of the selected SSRs for future varietal identification, the minimum number of SSR markers required to distinguish each of the mini-core accessions, at one or more loci, was selected using MinimalMarker software (Fujii et al., 2013).

2.5 | Selection of the mini-core collection and its genetic features

The number of accessions to be included in the mini-core collection was determined in such a way that >90% of the total number of alleles detected in the base collection were retained. The percentage of alleles retained was calculated by repeated trials with each sample number. To estimate the genetic relationship among the 447 accessions in the base collection, a dendrogram was generated based on SSR data with MEGA 5.1 software using the pairwise genetic distance matrix and neighbor-joining method (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013). The selection of accessions from the base collection was based on the diversity represented in the cluster while ensuring that accessions from each cluster were represented in the mini-core collection. When clusters had multiple candidate accessions, utmost care was taken to select those accessions that had good seed tubers available at the IITA GRC for further planting and evaluation. To determine the level of genetic diversity retained in the mini-core collection, the fixation index (F_{ST}) of each locus was calculated using GeneAlEx 6.5 software. Analysis of molecular variance (AMOVA) of the base collection and the mini-core collection was also conducted using Arlequin version 3.5 software (Excoffier & Lischer, 2010).

Additionally, structure analysis was applied to reveal the population structure of the selected mini-core collection.

2.6 | Ploidy analysis

Ploidy variation is an important cytogenetic trait in the white Guinea yam that is associated with several phenotypic traits. Ploidy analysis of the base collection was carried out by flow cytometry using a Partec ploidy analyzer (Sysmex Partec). Rice was used as an internal standard for calibration of the measurements. Fully developed young leaves (approximately 5 mm × 5 mm) sampled from field-grown plants in 2012 were chopped using a razor blade in 0.4 ml nuclear extraction buffer (solution A of a high-resolution kit; Sysmex Partec). 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 staining for 5 min at room temperature (~25 ˚C), flow cytometry was conducted at a rate of 5–20 nuclei per second. The DNA index (DI) of each accession was calculated on the basis of the relative amount of DNA in nuclei at the G1 stage compared with that of the internal standard. 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, TDr-1673, for which the chromosome number was confirmed microscopically through chromosome counting as $2n = 40$.

2.7 | Morphological diversity in the selected mini-core collection

A total of 21 morphological traits showing clear variation among *D. rotundata* germplasm were collected following descriptors for yam (IPGRI/IITA, 1997) in IITA GRC at the field gene bank in 2011 and were used for analyses to verify morphological diversity retained in the selected mini-core collection. Correspondence analysis was applied to the 21 selected traits using BellCurve software of Excel (Social Survey Research Information Company). As a diversity index, the Shannon–Weaver diversity index [$H' = -\sum_{i=1}^{n} p_i \ln(p_i)$, where n and p_i represent the number of classes of a trait and the proportion of accessions in the ith class of a trait, respectively) was calculated to assess phenotypic diversity (Shannon & Weaver, 1948).

2.8 | Evaluation of the agronomic performance of the mini-core collection

To evaluate the agronomic performance of the mini-core collection, the selected mini-core accessions were multiplied at IITA from single seed tubers obtained from the IITA GRC. Preliminary evaluation of the mini-core accessions, together with the reference lines, was conducted at IITA during 2013 and 2014 growing seasons. Seed sets (approximately 100 g each) of each accession or reference line were planted in small plastic pots in May 2013 and 2014, after which sprouted seed sets were transplanted in the field. Three plants per accession or reference line were planted in each plot, with 1-m spacing both between rows and between plants within a row. The plots were arranged according to a randomized block design, with a single replication in 2013 and two replications in 2014. The following agronomic traits were recorded: (a) number of stems per plant, (b) growing period (days from sprouting to senescence of the entire aerial part), (c) number of tubers per plant, (d) total yield per plant, and (e) average tuber weight. To determine statistical differences for agronomic traits among the mini-core accessions, average values per plot were used and an analysis of variance (ANOVA) was conducted with three replications.

3 | RESULTS AND DISCUSSION

3.1 | SSR polymorphisms in the base collection

Of the 90 SSR markers used in this study, 16 were selected following a preliminary screening based on their reproducibility, polymorphisms, and ease of scoring (Table 1). These 16 markers were used to assess 447 *D. rotundata* accessions of the base collection. The N_a detected per locus ranged from 2 to 12, with a total of 96 alleles detected (Table 2). The H_o at each locus ranged from 0.000 to 0.946 (mean = 0.373), and H_e ranged from 0.192 to 0.884 (mean = 0.583) (Table 2). The PIC of each locus ranged from 0.183 to 0.874, with a mean of 0.549 (Table 2). In the structure analysis, ΔK showed a maximum at $K = 8$, suggesting that eight genetically distinct clusters were observed among accessions of the base collection (Figure 1). All triploid ($2n = 3x = 60$) accessions were grouped together in one cluster (Cluster D in Figure 1).

3.2 | Selection of the mini-core collection

To select the mini-core collection, a smaller number of accessions that together captured >90% of total allelic diversity were chosen (Figure 2). Accordingly, a total of 102 accessions were selected from the base collection (Supplemental Table S1), evenly from all the clusters in the structure analysis (Supplemental Figure S1). The selected mini-core collection contained 97.92% (i.e., 94 of 96 alleles) of the alleles detected in the base collection (Table 2). Also, H_o, H_e, and PIC of the mini-core collection and base collection

TABLE 1 List of the 16 simple sequence repeat (SSR) markers used for genetic diversity analysis and their corresponding repeat motif, sequence, melting temperature (T_m) guanine-cytosine (GC) content, and expected product size in base pairs

Locus	Repeat motif	Primer	Sequence (5′–3′)	T_m	GC	Expected product size[a]
				˚C	%	bp
YM7	(CTT)15	Forward	AGCATTGGGTCCTTTCATCC	58.62	50.00	203
		Reverse	ACAATTCACACAAAGCATGGC	58.99	42.86	
YM8	(AG)24	Forward	TCTTAGGCTTTGGGCAGGG	60.08	57.89	166
		Reverse	AGTATGCCTACCCTGTTCTTC	57.57	47.62	
YM16	(CT)13	Forward	TGAAGAGAATGTTGAGATCGTACC	59.11	41.67	150
		Reverse	TATCCGGCCCTCTCATTGG	59.31	57.89	
YM18	(GT)19	Forward	GACATTGGGGATCTCTTATCAT	56.97	40.91	266
		Reverse	TAGCAGCAGTAACGTTAAGGAA	57.04	40.91	
YM25	(AG)30	Forward	GATGGAGATGAGGAGGCCG	60.08	63.16	237
		Reverse	TTCGAAGCCAGAGCAAGTG	58.82	52.63	
YM27	(GTT)8	Forward	TCCAGCTCTTTAGCACAGG	57.51	52.63	231
		Reverse	AGGAGCATAGGCAACAAGC	58.57	52.63	
YM28	(CTT)8…(CTT)14	Forward	CCATTCCTATTTAAGTTCCCCT	57.64	40.91	333
		Reverse	GATGAAGAAGAAGGTGATGATG	56.31	40.91	
YM30	(GT)16	Forward	CCACAACTAAAAACACATGGAC	57.07	40.91	212
		Reverse	GTGGTAGGGTGTGTAGCTTCTT	57.43	50.00	
YM31	(AAG)9	Forward	AAGCCTAGTCGATGGGTGG	59.55	57.89	221
		Reverse	TGCTGTTCCAACTTCCAAGC	59.72	50.00	
YM43	(AAG)9(GA)7	Forward	GCCTTGTTTTGTTGATGCTTCG	60.27	45.45	178
		Reverse	CCAGCCCACTAATCCCTCC	59.93	63.16	
YM44	(AG)20	Forward	CGCAACCAGCAAAGGATTTA	61.12	45.00	156
		Reverse	ATTCTGTCTCTCAAAACCCCT	56.85	42.86	
YM49	(AG)26	Forward	TGGGGTGAGAGAGTAAGTGG	58.83	55.00	163
		Reverse	TCACCGGGGATCTTCTTGC	60.16	57.89	
YM50	(CTT)9	Forward	TTGCCCTTGGGATGTAGGG	59.77	57.89	234
		Reverse	CATCCCCGTTGTATCCTGC	58.74	57.89	
YM54	(CT)11…(AC)16	Forward	CACTTGCTCTCTCATCGGC	59.06	57.89	162
		Reverse	TTGACAACCTCTATTTTGCCC	57.18	42.86	
YM61	(GTT)26	Forward	AGTGGTGCTGTAGTAACTGGAA	56.58	45.45	252
		Reverse	CATGACTACCTTTCCTCAATCA	26.82	40.91	
YM69	(CT)6…(AGTT)5	Forward	CTCTCTACCTCCCAACAAAAAC	57.04	45.45	229
		Reverse	AATCTTGCACCACCTTTTCTAC	57.43	40.91	

[a]Tamiru et al. (2015).

were similar (Table 2). Mean F_{ST} was low (0.0012), with a range of 0.000 to 0.004 (Table 2), suggesting the both the base and selected mini-core collections were genetically highly diverse. There was no significant difference in genetic variation between the base collection and selected mini-core collection as revealed by AMOVA (Table 3). Taken together, our results revealed that the selected mini-core collection retained the genetic diversity present in the base collection.

3.3 | Utilization of SSR markers for germplasm identification

To select a set of SSR markers that could discriminate accessions in the mini-core collection, 102 mini-core accessions were analyzed with different SSR marker combinations using MinimalMarker software. Three sets of six SSR markers ([YM7, YM8, YM25, YM27, YM31, YM54], [YM8, YM18, YM25, YM27, YM31, YM54], and [YM8, YM25, YM27,

TABLE 2 Summary statistics on the 16 simple sequence repeat (SSR) markers used to assess the genetic diversity in the *Dioscorea rotundata* base collection ($n = 447$) and the selected mini-core collection ($n = 102$)

	N_a[a]		H_o[b]		H_e[c]		PIC[d]		
Locus	**Base**[f]	**Mini-core**	**Base**[f]	**Mini-core**	**Base**[f]	**Mini-core**	**Base**[f]	**Mini-core**	F_{ST}[e]
YM7	8	8	0.642	0.647	0.759	0.721	0.725	0.683	0.0031
YM8	9	9	0.000	0.000	0.882	0.875	0.870	0.862	0.0014
YM16	5	5	0.152	0.147	0.308	0.377	0.295	0.358	0.0008
YM18	9	9	0.161	0.176	0.809	0.805	0.787	0.782	0.0007
YM25	12	11	0.362	0.490	0.749	0.768	0.716	0.743	0.0027
YM27	4	4	0.468	0.461	0.483	0.458	0.449	0.426	0.0005
YM28	7	6	0.472	0.520	0.665	0.664	0.609	0.609	0.0009
YM30	2	2	0.353	0.382	0.295	0.309	0.253	0.261	0.0003
YM31	9	9	0.488	0.490	0.862	0.850	0.847	0.833	0.0007
YM43	3	3	0.148	0.147	0.360	0.365	0.319	0.314	0.0006
YM44	6	6	0.405	0.333	0.604	0.513	0.569	0.485	0.0041
YM49	2	2	0.094	0.098	0.244	0.093	0.229	0.089	0.0000
YM50	2	2	0.369	0.373	0.500	0.498	0.375	0.374	0.0004
YM54	12	12	0.946	0.990	0.884	0.890	0.874	0.880	0.0006
YM61	4	4	0.792	0.775	0.731	0.730	0.681	0.680	0.0004
YM69	2	2	0.123	0.098	0.192	0.093	0.183	0.089	0.0011
Overall	96	94	0.373	0.383	0.583	0.563	0.549	0.529	0.0012

[a]N_a, number of alleles.
[b]H_o, observed heterozygosity.
[c]H_e, expected heterozygosity.
[d]PIC, polymorphic information content.
[e]F_{ST}, fixation index.
[f]Base, base collection.

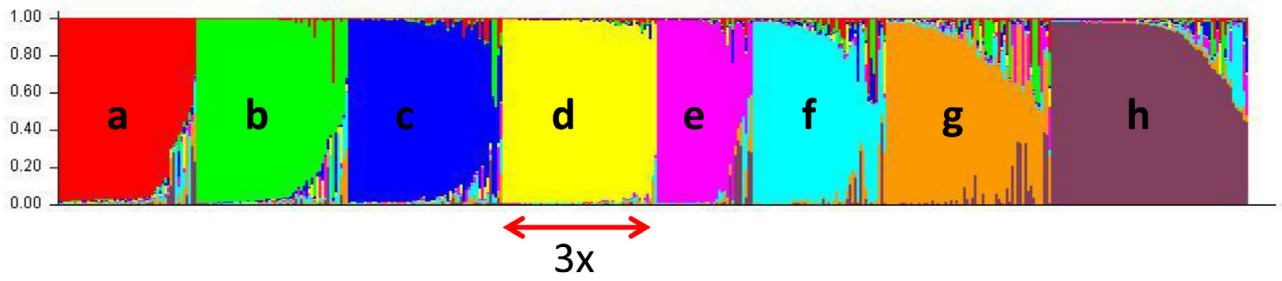

FIGURE 1 Genetic structure of *Dioscorea rotundata* base collection ($n = 447$) revealed by structure analysis with K (number of inferred ancestral populations) = 8 and sorted by expected populations (a–h). 3×, triploid ($2n = 3x = 60$)

YM31, YM54, YM61]) were subsequently found to distinguish all the 102 accessions. These results confirmed that the selected mini-core collection represents genetically unique (nonredundant at DNA level) accessions and suggest that a combination of a small number of SSRs can be used to carry out genotype identification in *D. rotundata*. These markers are important to facilitate utilization of the available genetic resources in yam breeding programs. Reliable identification of accessions or genotypes is crucial to ensure quality and uniformity of planting materials and authenticate progenies obtained from crosses. Accurate variety identification also aids in enforcing breeders' rights.

TABLE 3 Analysis of molecular variance (AMOVA) of the *Dioscorea rotundata* base collection ($n = 447$) and mini-core collection ($n = 102$)

Source of variation	df	SS[a]	Variance components	Variation	*P* value
				%	
Between populations	1	3.464	$-0.00108V_a$[b]	0.0	0.823ns[†]
Within populations	1,096	4,188.472	$3.82160Vb$[b]	100.0	
Total	1,097	4,191.936	3.82052		

[a]SS, sum of squares.
[b]V_a and V_b are the associated covariance components.
[†]ns, not significant at $\alpha = .05$ after 10,000 permutations.

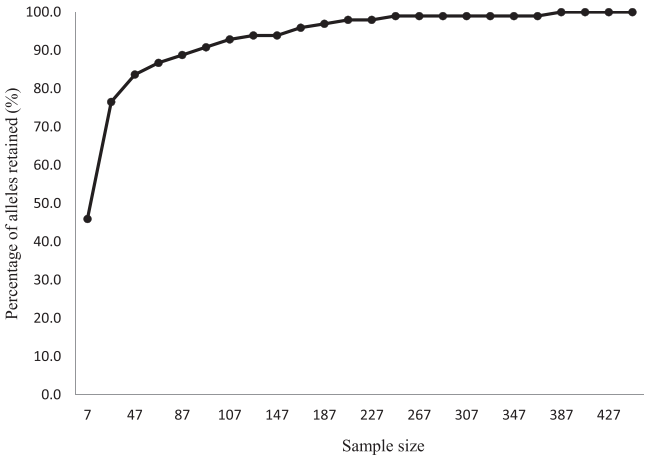

FIGURE 2 Percentage of simple sequence repeat (SSR) alleles retained in different sample sizes of *Dioscorea rotundata* base collection accessions ($n = 447$)

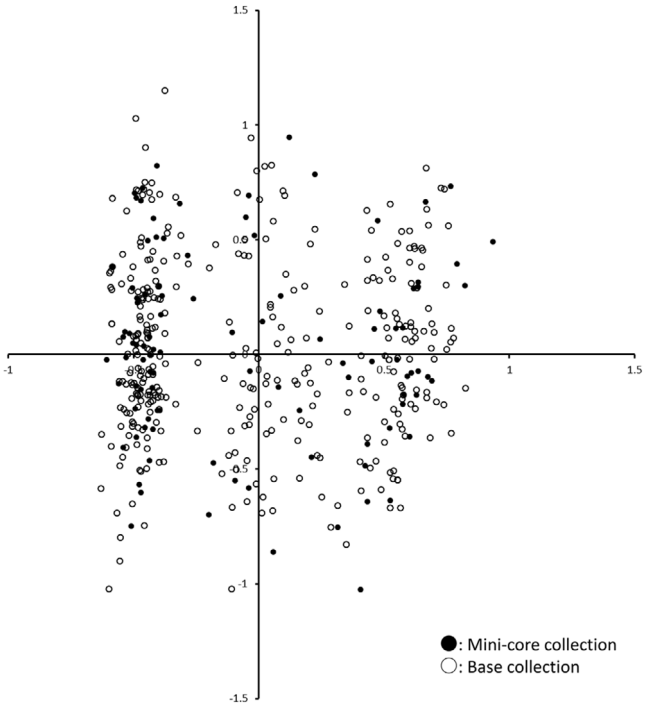

FIGURE 3 Scatterplot based on correspondence analysis of 21 morphological traits in the *Dioscorea rotundata* base collection ($n = 447$) and mini-core collection ($n = 102$)

3.4 | Morphological variability retained in the selected mini-core collection

The Shannon–Weaver diversity index (H') of the base collection for 21 morphological traits ranged from 0.25 (roots on the tuber surface) to 2.26 (stem color), and those of the mini-core collection ranged from 0.36 to 2.20, respectively (Table 4). The mean H' of the base collection and mini-core was 1.11 and 1.14, respectively. Of the 21 categorical traits analyzed, 14 showed similar or higher H' values in the mini-core collection compared with the base collection (Table 4). Furthermore, the Spearman's rank correlation (r) between base collection and mini-core collection was 0.964 ($P < .001$), suggesting the two collections were highly correlated with regard to morphological diversity. The first two axes of the scatter plot based on correspondence analysis of the 21 traits of the base collection accounted for 21.63 and 12.75% of the total variation, respectively, and the mini-core accessions were evenly distributed throughout the plot (Figure 3).

3.5 | Variations in the ploidy level and geographical origin of the base collection and mini-core collection

Of the 447 accessions analyzed for ploidy level, it was difficult to accurately estimate the ploidy levels of 13 accessions because of their broad flow cytometry histogram peaks. Therefore, these accessions were excluded and the ploidy information of the remaining 434 accessions was used. The DI of the standard diploid accession, TDr-1673, which has a chromosome number of $2n = 40$, was confirmed as 1.91. The base collection was subsequently divided into two groups based on DI values. Values for the first group ranged from 1.74 to 2.02, indicating a ploidy level similar to

TABLE 4 Shannon–Weaver diversity indices of morphological traits across the *Dioscorea rotundata* base collection ($n = 447$) and mini-core collection ($n = 102$)

Morphological trait	Base collection	Mini-core
Stem color	2.26	2.20
Absence or presence of waxiness	0.79	0.82
Absence or presence of spines	0.46	0.49
Spine shape	1.46	1.55
Leaf density	1.57	1.55
Leaf color	0.91	0.80
Leaf shape	1.52	1.48
Distance between lobes	1.33	1.32
Upward folding of the leaf along the main vein	1.00	1.00
Downward arching of the leaf along the main vein	1.00	1.00
Position of the widest part of the leaf	0.72	0.77
Sex	1.63	1.55
Inflorescence type	1.40	1.41
Spininess of roots	1.04	1.07
Tuber shape	1.52	1.62
Spiny roots on the tuber surface	0.77	0.73
Roots on the tuber surface	0.25	0.36
Prickly appearance of the tuber	0.98	1.00
Tuber flesh color: upper	1.16	1.26
Tuber flesh color: middle	0.94	1.03
Tuber flesh color: lower	0.68	0.89

that of TDr-1673. Meanwhile, the DI values of the second group ranged from 2.48 and 2.77—that is, a 1.3- to 1.5-fold increase compared with TDr-1673, suggesting triploidy ($2n = 3x = 60$). Of the 434 accessions analyzed, 381 (87.79%) were diploid ($2n = 40$), and the remaining 53 (12.21%) were triploid ($2n = 60$). Of the 102 accessions selected for the mini-core collection, 92 (90.2%) and 10 (9.8%) were diploid and triploid, respectively. In the population structure analysis, the diploid accessions were distributed across all clusters, whereas all triploid accessions were limited to cluster D (Figure 1).

Intraspecific ploidy variation is well known in the major cultivated yam species, including *D. rotundata*. Previously, two basic chromosome numbers ($x = 10$ and 20) were reported for *D. rotundata* (Dansi, Mignouna, Pillay, & Zok, 2001; Scarcelli, Dainou, Agbangla, Tostain, & Pham, 2005). In the present study, we adopted the basic chromosome number of 20, which was proposed on the basis of the segregation patterns of isozyme loci and microsatellite markers (Scarcelli et al., 2005). The presence of both diploid accessions ($2n = 40$) and triploid accessions ($2n = 60$) in the *D. rotundata* base collection is consistent with previous reports that documented the ploidy levels of *D. rotundata* and *D. cayenensis* (Dansi et al., 2001; Gamiette, Bakry, & Ano, 1999; Obidiegwu et al., 2009). Both the base collection and mini-core collection contained a similar proportion of diploids and triploids (Supplemental Table S1), confirming the base collection was adequately represented in the mini-core collection with respect to the ploidy level. Interestingly, structure analyses of the base and mini-core accessions revealed distinct clustering of the triploid accessions (Figures 1 and 4). This finding is consistent with a previous report that suggested an allo-polyploid origin of triploid *D. rotundata* accessions that harbored alleles of a wild-type species (Girma et al., 2014).

The selected mini-core collection was further validated for its representation of the base collection based on geographical origins of the accessions. As shown in Table 5, the mini-core collection also captured a proportional number of accessions representing different geographical origins similar to the base collection. Most of the accessions in the base population were from Nigeria (40.9%) and Togo (38.5%). Similarly, 84.3% of accessions in the mini-core collection were from Nigeria and Togo. Those from Benin (7.8%), Cote d'Ivoire (2.9%), and Ghana (3.9%) only had a limited representation in the mini-core collection. This imbalance in the geographical origin of the accessions in the base population and mini-core collection suggests the need to collect, characterize, and preserve additional accessions from the less represented but major yam-growing countries in West and Central Africa.

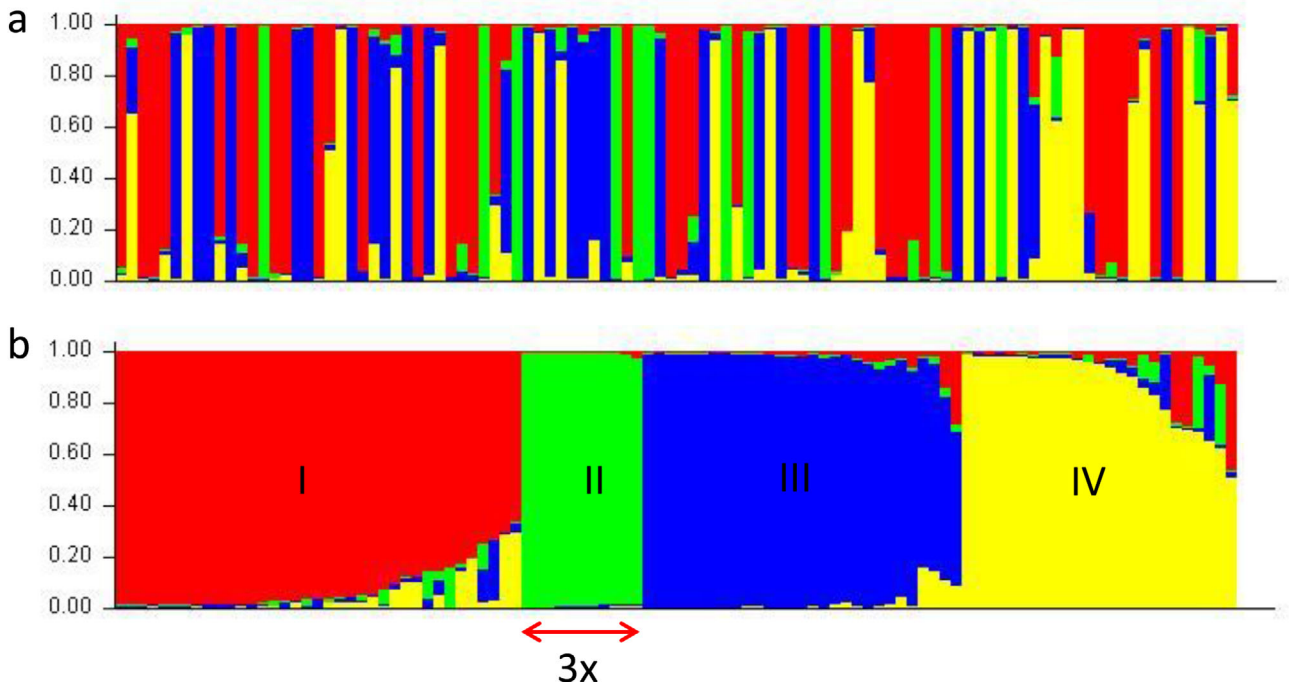

FIGURE 4 Genetic structure of *Dioscorea rotundata* mini-core collection ($n = 102$) revealed using the Structure version 2.3.4 software with K (number of inferred ancestral populations) = 4. (a) Bar plot in numerical order of accession number. (b) Sorted by expected populations (I–IV). 3×, triploid ($2n = 3x = 60$)

TABLE 5 Geographical origin of the accessions in the *Dioscorea rotundata* base collection ($n = 447$) and mini-core collection ($n = 102$)

Origin	No. of accessions (%)	
	Base collection	Mini-core
Benin	36 (8.1)	8 (7.8)
Burkina Faso	1 (0.2)	0 (0)
Cote d'Ivoire	21 (4.7)	3 (2.9)
Ghana	26 (5.8)	4 (3.9)
Guinea	7 (1.6)	1 (1.0)
Nigeria	183 (40.9)	32 (31.4)
Sierra Leone	1 (0.2)	0 (0)
Togo	172 (38.5)	54 (52.9)
Total	447	102

3.6 | Population structure of the selected mini-core collection

The analysis of the population structure of the mini-core collection revealed that 102 accessions mostly represented distinct genotypes; each accession is represented by a colored bar in Figure 4a. The ΔK was maximum at $K = 4$, suggesting four genetically distinct clusters among the mini-core accessions (Figure 4b). Also, all clusters included accessions from various geographic origins; thus, accessions in the mini-core collection were selected impartially with respect to geography and genetic diversity. With regard to the ploidy, Cluster II included all triploid accessions, indicating the genetic differentiation between diploid and triploid at the molecular level. This result was consistent with the result obtained for the base collection, as described above.

3.7 | Agronomic trait diversity among the mini-core accessions

The results from the field trials for agronomic traits of the mini-core collection conducted in 2013 and 2014 are summarized in Table 6. The ANOVA revealed highly significant differences ($P = .01$) among accessions for the following five traits: number of stems (mean = 2.15), growth period (mean = 161.6 days), number of tubers per plant (mean = 1.82), total yield per plant (mean = 1,351.6 g), and average tuber weight (738.2 g). Stable variation across three replications (for both years) was also observed in growth period, resulting in a low coefficient of variance (4.8%).

Increasing yam productivity in West Africa through breeding requires efficient phenotyping and genotyping techniques to identify materials with diverse target traits, such as high yield, early maturity, and optimal tuber size and shape. Because of its large plant size and long growing period, phenotyping for the analysis of the genetic basis of target traits is not an easy task in yam. In this study, we evaluated the mini-core collection together with the existing yam

TABLE 6 Variation in five agronomic traits in the *Dioscorea rotundata* mini-core collection ($n = 102$)

Trait	Mean	SD	CV	Summary of ANOVA	
				Variance	*F* value
No. of stems	2.15	0.747	0.347	4.82**	4.55
Growth period, d[a]	161.6	7.771	0.048	263.09**	3.18
No. of tubers per plant	1.82	1.115	0.620	2.48**	3.96
Total yield per plant, g	1,351.6	533.421	0.396	600,283.66**	4.31
Avg. tuber weight, g	738.2	492.022	0.489	485,691.17**	3.60

[a]Growth period, days from sprouting to senescence of the entire aerial part.
**Significant at the .01 probability level.

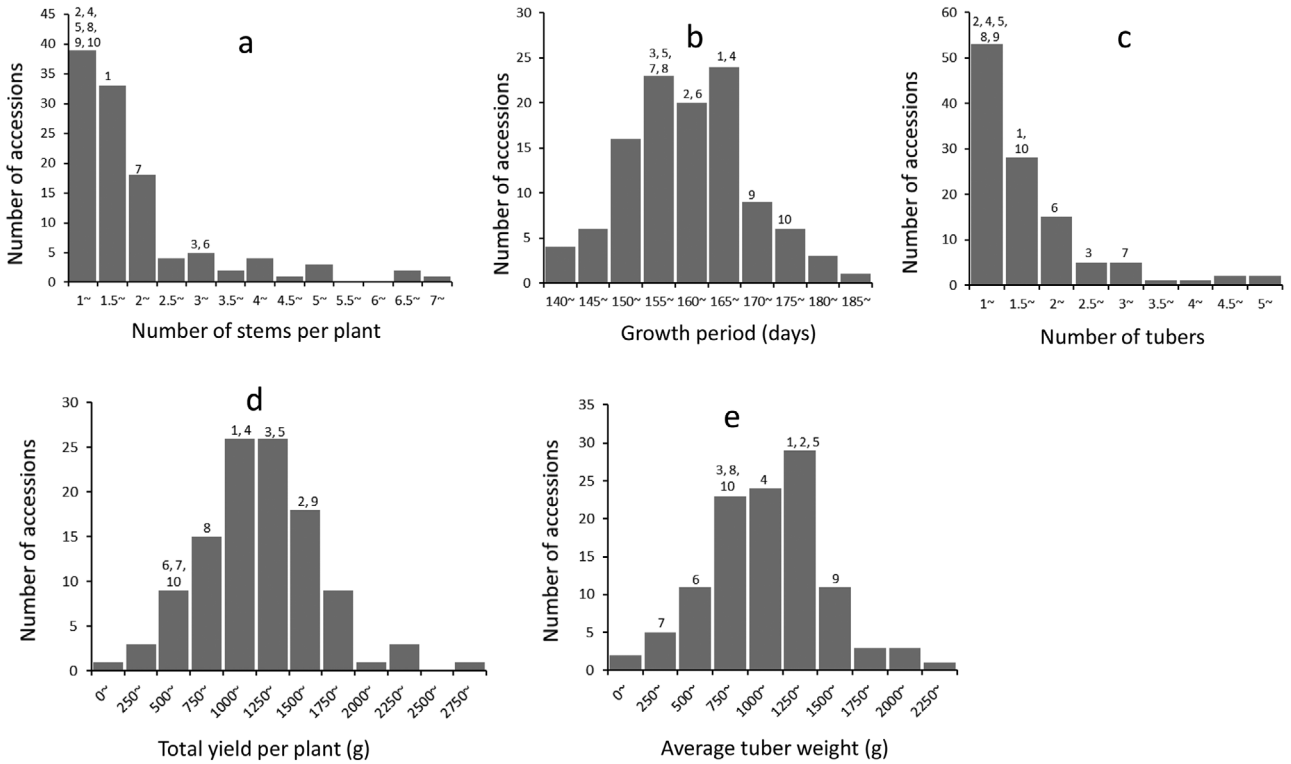

FIGURE 5 Variation in agronomic traits in the *Dioscorea rotundata* mini-core collection ($n = 102$). Distribution of the reference lines is shown with numbers from 1 to 10 on the histograms (IITA breeding lines: 1, TDr 97/00793; 2, TDr 97/00917; 3, TDr 97/00777; 4, TDr 99/02607; 5, TDr 95/01932 and local cultivars/landraces: 6, TDr EHOBIA; 7, TDr EHURU; 8, TDr OMI EFUN; 9, TDr POUNA; 10, TDr 04-219)

reference materials (breeding lines and landraces). Comparisons of the ranges of the five agronomic trait values revealed that the mini-core accessions were more diverse than the reference materials (Figure 5). Notably, the variation in the growth period, which had a coefficient of variation of 4.8%, was 155–176 d and 141–189 d for the reference lines and mini-core accessions, respectively. Although the selected mini-core collection does not cover the entire genetic diversity of *D. rotundata* in West Africa, it still contains considerable diversity that can be targeted for further studies aimed at dissecting the genetic basis of key traits and the genetic improvement of yams.

3.8 | Utilization of *D. rotundata* mini-core collection

Success in any breeding program depends largely on an efficient evaluation of existing germplasm to identify useful variation in target traits. However, germplasm evaluation is a challenging task when one deals with a large number of accessions with limited resources, which restricts proper utilization of germplasm (Upadhyaya, Gowda, Pundir, & Ntare, 2007). This is particularly the case for regionally important but underresearched local crops, such as yam in West and Central Africa. The mini-core collection developed for *D. rotundata*

in the present study represents a small, easy-to-use resource that is invaluable to accelerate genetic and genomic studies and improvement of this important crop. The whole genome sequence of *D. rotundata* is now available (Tamiru et al., 2017). This should allow the applications of several next-generation sequencing-based technologies to rapidly identify the genetic basis of important traits. Resequencing of the mini-core accessions and bulk sequencing of individuals pooled from segregating progeny can be used to identify candidate genes via genome-wide association studies and bulked segregant analysis, respectively (Huang et al., 2010; Takagi et al., 2013). This will further facilitate the development of DNA markers for marker-assisted selection, accelerating the deployment of useful alleles in yam breeding programs in West Africa. The SSRs identified in this study could also play a vital role in varietal identification in white Guinea yam and be routinely used in laboratories across West Africa, which often lack sophisticated facilities.

ACKNOWLEDGMENTS

This study was conducted under the project "Use of genomic information and molecular tools for yam germplasm utilization and improvement for West Africa (EDITS-Yam; 2011–2015)" of the Japan International Research Center for Agricultural Sciences, in collaboration with the Iwate Biotechnology Research Center and the International Institute of Tropical Agriculture. This work was also partly supported by the Japan Society for the Promotion of Science.

AUTHOR CONTRIBUTIONS

Babil Pachakkil: Investigation; Writing-original draft. Shinsuke Yamanaka: Conceptualization; Investigation; Writing-review & editing. Gezahegn Girma: Methodology; Validation. Ryo Matsumoto: Investigation; Validation. Muluneh Tamiru: Methodology; Validation. Ranjana Bhattacharjee: Validation. Michael Abberton: Resources; Validation. Satoru Muranaka: Validation. Robert Asiedu: Supervision. Hiroko Takagi: Conceptualization; Project administration; Supervision.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ORCID

Shinsuke Yamanaka https://orcid.org/0000-0003-3836-6385
Satoru Muranaka https://orcid.org/0000-0002-2810-1247

REFERENCES

Asiedu, R., Ng, S. Y. C., Bai, K. V., Ekanayake, I. J., & Wanyera, N. M. W. (1998). Genetic improvement. In G. C. Orkwor, R. Asiedu, & I. J. Ekanayake (Eds.), *Food yams: Advances in research* (pp. 63–104). Ibadan, Nigeria: IITA and National Root Crops Research Institute.

Dansi, A., Dantsey-Barry, H., Dossou-Aminon, I., N'Kpenu, E. K., Agre, A. P., Sunu, Y. D., … Vodouhe, R. (2013). Varietal diversity and genetic erosion of cultivated yams (*Dioscorea cayenensis* Poir–*D. rotundata* Lam complex and *D. alata* L.) in Togo. *International Journal of Biodiversity and Conservation*, *5*, 223–239.

Dansi, A., Mignouna, H. D., Pillay, M., & Zok, S. (2001). Ploidy variation in the cultivated yams (*Dioscorea cayenensis–Dioscorea rotundata* complex) from Cameroon as determined by flow cytometry. *Euphytica*, *119*, 301–307. https://doi.org/10.1023/A:1017510402681

Dansi, A., Mignouna, H. D., Zoundjihekpon, J., Sangare, A., Ahoussou, N., & Asiedu, R. (2000). Identification of some Benin Republic's Guinea yam (*Dioscorea cayenensis/Dioscorea rotundata* complex) cultivars using randomly amplified polymorphic DNA. *Genetic Resources and Crop Evolution*, *47*, 619–625. https://doi.org/10.1023/A:1026589702426

Dansi, A., Mignouna, H. D., Zoundjihekpon, J., Sangare, A., Asiedu, A., & Quin, F. M. (1999). Morphological diversity, cultivar groups and possible descent in the cultivated yams (*Dioscorea cayenensis/D. rotundata*) complex in Benin Republic. *Genetic Resources and Crop Evolution*, *46*, 371–388. https://doi.org/10.1023/A:1008698123887

Demuyakor, B., Dukrog, T. M., & Chikpah, S. K. (2013). Yam germplasm in Ghana: A survey on storage and varietal properties of *Dioscorea rotundata–alata* in northern region of Ghana. *International Journal of Scientific & Technology Research*, *2*, 170–175.

Ebana, K., Kojima, Y., Fukuoka, S., Nagamine, T., & Kawase, M. (2008). Development of mini core collection of Japanese rice landrace. *Breeding Science*, *58*, 281–291. https://doi.org/10.1270/jsbbs.58.281

Evanno, S., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

FAO. (2018). *FAOSTAT database*. Rome: FAO. Retrieved from https://www.fao.org/faostat/en/#data

Frankel, O. H. (1984). Genetic perspectives of germplasm conservation. In W. K. Arber, et al. (Eds.), *Genetic manipulation: Impact on man and society* (pp. 161–170). Cambridge, UK: Cambridge University Press.

Fujii, H., Ogata, T., Shimada, T., Endo, T., Iketani, H., Shimizu, T., … Omura, M. (2013). MINIMAL MARKER: An algorithm and computer program for the identification of minimal sets of discriminating DNA markers for efficient variety identification. *Journal of Bioinfomatics and Computational Biology*, *11*(2). https://doi.org/10.1142/S0219720012500229

Fulton, T. M., Chunwongse, J., & Tanksley, S. D. (1995). Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Molecular Biology Reporter*, *13*, 207–209. https://doi.org/10.1007/BF02670897

Gamiette, F., Bakry, F., & Ano, G. (1999). Ploidy determination of some yam species (*Dioscorea* spp.) by flow cytometry and conventional chromosomes counting. *Genetic Resources and Crop Evolution*, *46*, 19–27. https://doi.org/10.1023/A:1008649130567

Girma, G., Bhattacharjee, R., Lopez-Montes, A., Gueye, B., Ofodile, S., Franco, J., & Abberton, A. (2018). Re-defining the yam (*Dioscorea* spp.) core collection using morphological

traits. *Plant Genetic Resources*, *16*, 193–200. https://doi.org/10.1017/S1479262117000144

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*, 1783–1794. https://doi.org/10.1007/s00122-014-2339-2

Girma, G., Natsume, S., Carluccio, A. V., Takagi, H., Matsumura, H., Uemura, A., … Tamiru, M. (2019). Identification of candidate flowering and sex genes in white Guinea yam (*D. rotundata* Poir.) by SuperSAGE transcriptome profiling. *PLOS ONE*, *14*(9). https://doi.org/10.1371/journal.pone.0216912

Hamon, P., & Bakary, T. (1990). Characterization of traditional yam varieties belonging to the *Dioscorea cayenensis–rotundata* complex by their isozymic patterns. *Euphytica*, *46*, 101–107. https://doi.org/10.1007/BF00022303

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., … Han, B. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics*, *42*, 961–967. https://doi.org/10.1038/ng.695

IPGRI/IITA. (1997). *Descriptors for yam (Dioscorea spp.)*. Rome: IPGRI/IITA.

Mahalakshmi, V., Ng, Q., Atalobhor, J., Ogunsola, D., Lawson, M., & Ortiz, R. (2007). Development of a West African yam *Dioscorea* spp. core collection. *Genetic Resources and Crop Evolution*, *54*, 1817–1825. https://doi.org/10.1007/s10722-006-9203-4

Mignouna, H., Mank, R., Ellis, T., van den Bosch, N., Asiedu, R., Ng, C., & Peleman, J. (2002). A genetic linkage map of Guinea yam (*Dioscorea rotundata* Poir.) based on AFLP markers. *Theoretical and Applied Genetics*, *105*, 716–725. https://doi.org/10.1007/s00122-002-0911-7

Mignouna, H. D., Abang, M. M., Wanyera, N. W., Chikaleke, V. A., Asiedu, R., & Thottappilly, G. (2005). PCR marker-based analysis of wild and cultivated yams (*Dioscorea* spp.) in Nigeria: Genetic relationships and implication for ex situ conservation. *Genetic Resources and Crop Evolution*, *52*, 755–763. https://doi.org/10.1007/s10722-004-6128-7

Mignouna, H. D., Dansi, A., & Zok, S. (2002). Morphological and isozymic diversity of the cultivated yams (*Dioscorea cayenensis/Dioscorea rotundata* complex) of Cameroon. *Genetic Resources and Crop Evolution*, *49*, 21–29. https://doi.org/10.1023/A:1013805813522

Obidiegwu, J., Loureiro, J., Ene-Obong, E., Rodriguez, E., Kolesnikova-Allen, M., Santos, C., … Asiedu, R. (2009). Ploidy level studies on the *Dioscorea cayenensis/Dioscorea rotundata* complex core set. *Euphytica*, *169*, 319–326. https://doi.org/10.1007/s10681-009-9960-1

Odong, T. L., Jansen, J., van Eeuwijk, F. A., & van Hintum, T. J. L. (2013). Quality of core collections for effective utilization of genetic resources review, discussion and interpretation. *Theoretical and Applied Genetics*, *126*, 289–305. https://doi.org/10.1007/s00122-012-1971-y

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, *28*, 2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Pritchard, J. K., Stephens, P., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959.

Sartie, A., Asiedu, R., & Franco, J. (2012). Genetic and phenotypic diversity in a germplasm working collection of cultivated tropical yams (*Dioscorea* spp.). *Genetic Resources and Crop Evolution*, *59*, 1753–1765. https://doi.org/10.1007/s10722-012-9797-7

Scarcelli, N., Dainou, O., Agbangla, C., Tostain, S., & Pham, J. L. (2005). Segregation patterns of isozyme loci and microsatellite markers show the diploidy of African yam *Dioscorea rotundata* ($2n = 40$). *Theoretical and Applied Genetics*, *111*, 226–232. https://doi.org/10.1007/s00122-005-2003-y

Shannon, C. E., & Weaver, W. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423 & 623–656.

Shehzad, T., Okuizumi, H., Kawase, M., & Okuno, K. (2009). Development of SSR-based sorghum (*Sorghum bicolor* (L.) Moench) diversity research set of germplasm and its evaluation by morphological traits. *Genetic Resources and Crop Evolution*, *56*, 809–827. https://doi.org/10.1007/s10722-008-9403-1

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., … Terauchi, R. (2013). QTL-seq: Rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant Journal*, *74*, 174–183. https://doi.org/10.1111/tpj.12105

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*, 1744–1753. https://doi.org/10.2135/cropsci2006.11.0719

Tamiru, M., Natsume, S., Takagi, H., White, B., Yageshi, H., Shimizu, 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*. https://doi.org/10.1186/s12915-017-0419-x

Tamiru, M., Yamanaka, S., Mitsuoka, C., Babil, P., Takagi, H., Lopez-Montes, A., … Terauchi, R. (2015). Development of genomic simple sequence repeat markers for yam. *Crop Science*, *55*, 2191–2200. https://doi.org/10.2135/cropsci2014.10.0725

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, *30*, 2725–2729. https://doi.org/10.1093/molbev/mst197

Upadhyaya, H. D., Gowda, C. L. L., Pundir, R. P. S., & Ntare, B. R. (2007). Use of core and mini core collections in preservation and utilization of genetic resources in crop improvement. In S. R. Vodouhe, et al. (Eds.), *Plant genetic resources and food security in West and Central Africa* (pp. 96–107). Rome: Bioversity International.

SUPPORTING INFORMATION

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

How to cite this article: Pachakkil B, Yamanaka S, Girma G, et al. Simple sequence repeat-based mini-core collection for white Guinea yam (*Dioscorea rotundata*) germplasm. *Crop Science*. 2021;1-12. https://doi.org/10.1002/csc2.20431