Theor Appl Genet (2014) 127:1783–1794
DOI 10.1007/s00122-014-2339-2

ORIGINAL PAPER

Next-generation sequencing based genotyping, cytometry and phenotyping for understanding diversity and evolution of guinea yams

Gezahegn Girma · Katie E. Hyma · Robert Asiedu · Sharon E. Mitchell · Melaku Gedil · Charles Spillane

Received: 10 March 2014 / Accepted: 22 May 2014 / Published online: 1 July 2014

Abstract
Key message **Genotyping by sequencing (GBS) is used to understand the origin and domestication of guinea yams, including the contribution of wild relatives and polyploidy events to the cultivated guinea yams.**
Abstract Patterns of genetic diversity within and between two cultivated guinea yams (*Dioscorea rotundata* and *D. cayenensis*) and five wild relatives (*D. praehensilis*, *D. mangenotiana*, *D. abyssinica*, *D. togoensis* and *D. burkilliana*) were investigated using next-generation sequencing (genotyping by sequencing, GBS). Additionally, the two cultivated species were assessed for intra-specific morphological and ploidy variation. In guinea yams, ploidy level is correlated with species identity. Using flow cytometry a single ploidy level was inferred across *D. cayenensis* (3x, $N = 21$), *D. praehensilis* (2x, $N = 7$), and *D. mangenotiana* (3x, $N = 5$) accessions, whereas both diploid and triploid (or aneuploid) accessions were present in *D. rotundata* ($N = 11$ and $N = 32$, respectively). Multi-dimensional scaling and maximum parsimony analyses of 2,215 SNPs revealed that wild guinea yam populations form discrete genetic groupings according to species. *D. togoensis* and *D. burkilliana* were most distant from the two cultivated yam species, whereas *D. abyssinica*, *D. mangenotiana*, and *D. praehensilis* were closest to cultivated yams. In contrast, cultivated species were genetically less clearly defined at the intra-specific level. While *D. cayenensis* formed a single genetic group, *D. rotundata* comprised three separate groups consisting of; (1) a set of diploid individuals genetically similar to *D. praehensilis*, (2) a set of diploid individuals genetically similar to *D. cayenensis*, and (3) a set of triploid individuals. The current study demonstrates the utility of GBS for assessing yam genomic diversity. Combined with morphological and biological data, GBS provides a powerful tool for testing hypotheses regarding the evolution, domestication and breeding of guinea yams.

Communicated by Ian D. Godwin.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-014-2339-2) contains supplementary material, which is available to authorized users.

G. Girma · R. Asiedu · M. Gedil (✉)
International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria
e-mail: m.gedil@cgiar.org

G. Girma · C. Spillane (✉)
Genetics and Biotechnology Lab, Plant and AgriBiosciences Research Centre (PABC), School of Natural Sciences, National University of Ireland Galway, University Road, Galway, Ireland
e-mail: charles.spillane@nuigalway.ie

K. E. Hyma · S. E. Mitchell
Institute for Genomic Diversity, Cornell University, Ithaca, NY, USA

K. E. Hyma
Bioinformatics Facility, Cornell University, Ithaca, NY, USA

Introduction

Yams (*Dioscorea* spp.) are an economically important edible tuber food crop in tropical regions worldwide. The use of yams for consumption and a source of income, combined with its socio-cultural importance (Tortoe et al. 2012), makes it one of the most important food and livelihood security crops in Africa. Yams are located within a large genus comprising approximately 450 species (Govaerts et al. 2007). The *Dioscorea* species are herbaceous climbing monocots within the family Dioscoreaceae (Coursey 1967). The poor flowering, seed production and

germination of cultivated yams (Lebot 2009; Mignouna et al. 2007) restricts farm-level production to clonal propagation (Scarcelli et al. 2013). However, true seeds are being used in breeding programs and few reports indicate the occurrence of natural hybridization (Cornet et al. 2010; Scarcelli et al. 2006b) indicating the possible propagation of wild yams through botanic seeds. While several yam species are present in West Africa, the native *D. rotundata* Poiret and *D. cayenensis* Lamarck (also referred to as Guinea yams or the *D. cayenensis–rotundata* complex) are the most important and most widely cultivated (Mignouna et al. 2003).

Guinea yams were likely domesticated by farmers from wild yams of the section *Enantiophyllum* (Burkill 1960; Terauchi et al. 1992; Zannou et al. 2006). Domestication is still ongoing in Benin (Scarcelli et al. 2006a, b; Zannou et al. 2006), although the practice is limited to a small number of farmers (Cornet et al. 2010). A recent study by (Zannou et al. 2006) indicated that Benin farmers consider the wild yam tubers to be edible after three consecutive plantings and harvests. As part of the domestication process, the farmers select wild forms for tuber shape and taste which resemble some cultivated varieties in their vegetative parts (Zannou et al. 2004). In addition, several authors have reported the direct use of wild yams as a source of food in West Africa (Bahuchet et al. 1991; Sato 2001).

Although Africa represents 96 % of the total production of yams worldwide (estimated at 40 million tonnes average for the period of 1992–2011), no African country is among the top five countries delivering the highest yields (FAOSTAT 2013). The top five countries producing high yield per area include Japan, Papua New Guinea, Tonga, Jamaica, and Portugal, whereas Nigeria, Cote d'Ivoire, Ghana, Benin and Togo are the top five countries in terms of total production.

African farmers face multiple constraints to achieve high yam output. Diseases and storage pests are the major constraints to yam production in West Africa and, over time, these limitations have become more severe (Aidoo et al. 2011; Amusa et al. 2003; Baimey et al. 2006). Breeding for improved varieties in yam is challenging due to the polyploid nature of the crop. Transfer of desirable genes from the secondary genepool of wild relatives to the cultivated primary genepool remains difficult in many crops, including in yams (Spillane and Gepts 2001). Yet, the wild relatives of yams can harbor desirable genes and genetic diversity that has potential for utilization in breeding efforts to enhance the agronomic performance of yam cultivars. Therefore, understanding the genetic relationship and the biology of yam wild relatives is important for improving cultivated yam species.

To date, there is no clear-cut information on the extent of genetic diversity within and between cultivated guinea yam species and their wild relatives. Genetic diversity in cultivated and wild guinea yams has been investigated using AFLPs, RAPDs, microsatellites and RFLPs (Ramser et al. 1997; Scarcelli et al. 2006b; Terauchi et al. 1992). However, these studies could not discriminate some of the wild species from cultivated types, and concluded that the wild and cultivated *Dioscorea* species were very closely related. A recent study on guinea yam collections from Ethiopia using SSR loci (Mengesha et al. 2013) found no clear distinction between cultivated and wild species.

Morphological characterization studies on cultivars from Benin and Cameroon distinguished individuals and further classified them into *D. rotundata*, *D. cayenensis*, and *D. rotundata* × *D. cayenensis* groups (Dansi et al. 1999; Mignouna et al. 2002). These and other authors suggested the possibility of natural hybridization between different species as a cause of cultivars with heterogeneous morphological traits. However, the difficulty to find reliable and stable morphological traits to discriminate between cultivars was also indicated. *D. abyssinica* Hochst. ex Kunth, *D. praehensilis* Benth, *D. burkilliana* J. Miege, *D. mangenotiana* J. Miege and *D. liebrechtsiana* De Wild were suggested as progenitors of cultivated guinea yam based on shared morphological similarity between plants of wild and cultivated species (Dansi et al. 1999; Mignouna et al. 2002; Terauchi et al. 1992).

Guinea yams, *D. rotundata* and *D. cayenensis*, are polyploid species in which different lines can display different ploidy levels. It has been proposed that *D. rotundata* is a tetraploid with a basic chromosome number of 10 ($x = 10$) (Dansi et al. 2001; Gamiette et al. 1999; Obidiegwu et al. 2009). Hexaploid and octaploid individuals have been reported in *D. cayenensis* based on DNA flow cytometry, using *Solanum lycopersium* L. (Obidiegwu et al. 2009) and the tetraploid *D. rotundata* (Dansi et al. 2001; Gamiette et al. 1999) as internal standards. However, a study based on segregation patterns of isozyme and microsatellite loci has indicated that *D. rotundata* is diploid, with a chromosome number of 20 ($2n = 40$) (Scarcelli et al. 2005). Flow cytometry histograms for *D. cayenensis–rotundata* were not distinct from those of its related wild relatives (*D. abyssinica*, *D. mangenotiana*, *D. burkilliana* and *D. praehensilis*) (Gamiette et al. 1999).

Similar ploidy studies have been performed for two of the other agriculturally most important *Dioscorea* species. *D. trifida* Linnaeus, once thought to be octoploid, is now considered to be an autotetraploid (Bousalem et al. 2006). Likewise, a study based on the microsatellite segregation analysis of four different progenies has demonstrated that *D. alata* Linnaeus accessions can be diploid, triploid and tetraploid ($2n$ = 2x, 3x, 4x), respectively, and not tetraploid, hexaploid and octoploid ($2n$ = 4x, 6x, 8x) as previously assumed, with a basic chromosome number of

20 (Arnau et al. 2009). A study by Nemorin et al. (2012) further confirmed the autotetraploid nature of the $2n = 80$ clones of *D. alata*. However, the extent of polyploidy is not yet known across guinea yam genepools, which represents an important knowledge gap in understanding the biology and agricultural performance of cultivated guinea yams. For instance, it is possible that ploidy could play an important role in both the morphological and agronomical characteristics of guinea yams.

Flow cytometric data alone cannot provide conclusive evidence of ploidy level. Emshwiller (2002) indicated that it can be difficult to distinguish DNA content levels among close ploidy levels. Previously reported as heptaploid ($2n = 7x = 49$), *Oxalis tuberosa* Molina, was later found to be actually octoploid ($2n = 8x = 64$) using a combination of flow cytometry and molecular evidence. This highlighted the importance of combining both molecular and cytological data in confirming ploidy levels.

Next-generation based genotyping procedures, such as genotyping by sequencing (GBS), represent high-marker density approaches which can help reveal the extent of genetic relatedness and genetic variation within and between cultivated species and their wild relatives (Spindel et al. 2013). The GBS approach is based on reducing genome complexity with restriction enzymes, coupled with multiplex NGS for high-density single nucleotide polymorphism (SNP) discovery (Elshire et al. 2011). The genome-wide molecular marker discovery, highly multiplexed genotyping, flexibility and low cost of GBS make it an excellent tool in plant genetics and breeding (Deschamps et al. 2012; Poland and Rife 2012). The development of a robust SNP calling pipeline, Universal Network Enabled Analysis Kit (UNEAK) (Lu et al. 2013) facilitates the use of GBS for genomic diversity and genetic relationship studies in species that lack a reference genome sequence, such as guinea yams.

GBS offers an advantage as it can simultaneously discover polymorphisms and obtain genotypic information across the population of interest. Poland and Rife (2012) have highlighted that it represents a fast and inexpensive approach that can enable genotyping of large populations of selection candidates within breeding programs. This can further assist breeders to more efficiently choose genetically diverse parents in breeding programs that employ both interspecific and intraspecific hybridization. GBS diversity assessment can also provide a means for identifying potential gaps in species collection and further guiding germplasm collecting missions. Taking advantage of the power of a GBS approach, this study aims (1) to increase understanding of genomic diversity and genetic structure of guinea yams and their wild relatives, and (2) to investigate the morphological and ploidy variation within and between cultivated guinea yam species.

Materials and methods

Plant materials

A total of seven guinea yam species were used for this study. All individual accessions of the two cultivated species, *D. rotundata* and *D. cayenensis*, including two of the wild species, *D. mangenotiana* and *D. praehensilis*, were obtained from IITA field genebank. The *D. togoensis* accessions were collected from the IITA forest, where they are conserved in situ. Accessions of two other wild species, *D. abyssinica* and *D. burkilliana* were kindly supplied by Professor Alexander Dansi from Benin. The accessions of *D. burkilliana* were collected from wild populations while *D. abyssinica* was collected from Northern region of Benin where there is evidence of ongoing domestication of wild yams by farmers (Scarcelli et al. 2006a, b; Zannou et al. 2006) (Fig. 1). All of the individual accessions (Supplementary Table 1) were used for genotyping; the two cultivated species (comprising 43 *D. rotundata* and 21 *D. cayenensis*) were also assessed for morphological variation. The cultivated species including two of the wild species, *D. mangenotiana* and *D. praehensilis* were evaluated for ploidy level using a flow cytometry approach.

Phenotyping of yam accessions

All individuals of the cultivated species within the IITA field genebank were assessed in 2012 for intra-specific and inter-specific morphological variation. The materials were planted following standard procedures (Dumet and Ogunsola 2008) as routine field genebank regeneration during the main growing season at the IITA experimental plot, Ibadan (latitude: 7°30′8″N; longitude: 3°54′38″E), Nigeria. Data were collected from three individuals planted and labeled as A, B and C per accession. Fourteen yam morphological descriptors (IPGRI/IITA 1997) were used. The descriptors consisted of stem color, vigor, presence and absence of barky patches and waxiness, leaf shape, leaf color, distance between lobes, sex, number of inflorescences, flower color, tuber flesh color (observed on the upper, middle and lower part), and tuber beneath skin color.

Ploidy analysis

The ploidy level of the cultivated species was analyzed using a flow cytometry approach. Two of the wild species, *D. mangenotiana* and *D. praehensilis*, whose genome size (611 Mbp) is similar to that of the cultivated *D. rotundata* (Hamon et al. 1992) were also analyzed using a diploid *D. rotundata* accession (TDr 1673, $2x = 2n = 40$) as standard. Ploidy analysis was performed using previously described protocols (Babil et al. 2010). Young leaves were

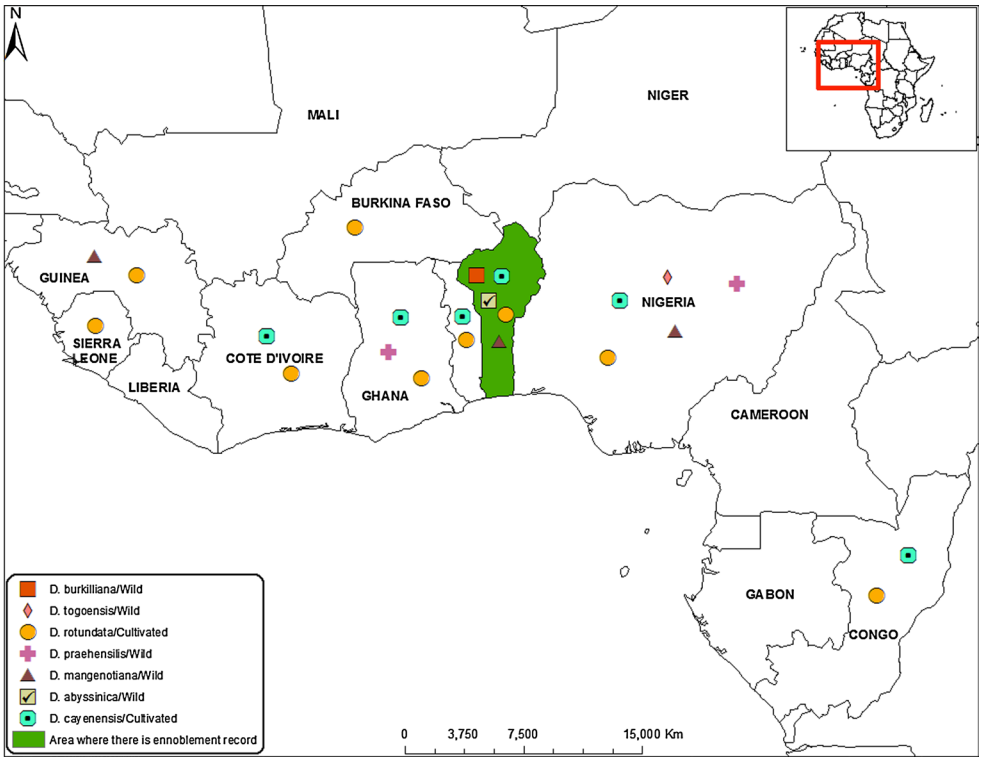

Fig. 1 Map indicating collection sites for wild and cultivated guinea yam species used in this study. Benin is *shaded* in *green* as this is the region where there is evidence of ongoing domestication of wild yams by farmers, via a farmer-driven selection process called ennoblement (color figure online)

collected from individual plants. A leaf blade of approximately 5 mm^2 was chopped to homogenize the tissue by adding 500 μL ice cold OTTO I buffer (0.1 M citric acid monohydrate 0.5 % Tween 20). The homogenate was filtered through a 50-μm-pore size nylon filter into a plastic tube. The cell suspension was incubated for 5 min at room temperature. The nuclear DNA was stained by adding 2 mL of OTTO II buffer (0.4 M Na_2PO_4 supplemented with 4 μg/mL of DAPI—4,6-diamidino-2-phenylindole) and 1 μL/mL mercaptoethanol to each tube. Relative fluorescence intensity was measured to determine the ploidy by using the standard as internal reference. The flow cytometer was adjusted so that the peak representing the G1 nuclei of the diploid standard (TDr 1673) was set at channel 50.

Yam DNA samples

A total of 95 yam accessions comprising the two cultivated species including five of its wild relatives were genotyped (Table 1; Fig. 1). Leaf samples were collected and lyophilized. DNA was extracted using a Qiagen-DNeasy plant mini kit (QIAGEN GmbH). Samples were quantified using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific). For further quality and quantity assessments, 1 μL (100 ng) DNA of all samples was run on 1 % w/v agarose gel along with 500 ng of two λ *Hin*dIII size standards per gel. A trial digestion was done for ten randomly selected DNA samples using 1 U of *Hin*dIII, which were run on a 1 % w/v agarose gel along with the λ *Hin*dIII size standards. Two different concentrations of DNA (100 and 500 ng) were used for each digest. The restriction enzyme digested better at the lower DNA concentration (100 ng).

Table 1 Summary of ploidy levels across different guinea yam species

Species	Other names	Number of individuals	Ploidy level
D. rotundata	White yam	32	2x
		11	3x
		1[a]	2x
D. cayenensis	Yellow yam	21	3x
D. abyssinica	Wild spp.	2	NA
D. burkilliana	Wild spp.	7	NA
D. mangenotiana	Wild spp.	5	3x
D. praehensilis	Wild spp.	7	2x
D. togoensis	Wild spp.	6	NA

NA ploidy level not determined by flow cytometry

[a] Individual used as standard for flow cytometry

GBS libraries and sequencing

GBS libraries were prepared and analyzed at the Institute for Genomic Diversity (IGD) at Cornell University, following (Elshire et al. 2011). PstI enzyme was used for digestion

and for creating a library containing 96 unique barcodes (95 uniquely named samples and one negative control containing no DNA). The GBS library was sequenced on a single Illumina HiSeq lane. A total of 118,383,523 100 bp reads were generated and used for SNP calling.

Phenotypic data analysis

Multiple correspondence analyses was performed for the categorical phenotypic data with FactoMineR package (Lê et al. 2008) using R software (R Core Team 2013) to detect the underlying pattern and structures in a data set.

Analysis of GBS data

A modified version of the non-reference GBS SNP calling pipeline UNEAK (http://www.maizegenetics.net/gbs-bioinformatics), as implemented in Tassel Version 3.0.160 (Lu et al. 2013), was used for SNP calling (see supplementary materials for XML configuration files and barcode keyfile). A total of 6,371 SNPs were identified. A filtered dataset was created using VCFtools version v0.1.10 (Danecek et al. 2011) by first filtering genotypes with quality scores less than 98 (−GQ 98), and then removing SNP loci with more than 90 % missing data (−geno 0.1). A total of 2,215 SNP loci remained after filtering. Raw data are deposited at the Dryad Digital Repository (https://datadryad.org/). See supplementary material for complete details of analysis methods used.

Multi-dimensional scaling analysis (MDS) was conducted using PLINK version v1.07 (Purcell et al. 2007). Results were used to assign individuals to seven distinct groups: (1) *D. burkilliana*, (2) *D. cayenensis*, (3) *D. togoensis*, (4) *D. rotundata* (2x), (5) *D. rotundata* (3x), (6) *D. praehensilis*, and (7) *D. mangenotiana* (Fig. 2). The individual db-8, originally identified as *D. burkilliana* based on morphology, was found to be a potential mis-identified *D. mangenotiana* based on its pattern of heterozygosity and genetic similarity and was treated as *D. mangenotiana* for further analyses.

Nucleotide distances (substitution rates per site) were calculated between and within groups using MEGA version 5 (Tamura et al. 2011). A maximum parsimony (MP) analysis was carried out on the 2,215 SNPs using PHYLIP. Five hundred sets of weights were generated by bootstrapping (seqboot). From these 500 replicates, 25,370 MP trees representing 23,588 topologies were generated (dnapars). Trees were re-rooted at the longest branch using Newick tools v.1.6 (Junier and Zdobnov 2010) and visualized with Densitree v2.1.10 (Bouckaert 2010). A cluster analysis using weighted correlation network was performed on genotypes in R (R Core Team 2013; R Development Core Team 2010) using the R package WGCNA (Langfelder and Horvath 2008).

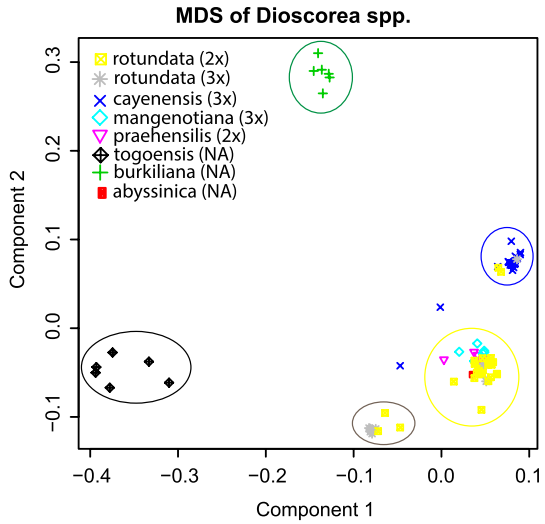

Fig. 2 Multi-dimensional scaling (MDS) analysis of yam GBS genotypes

The proportion of heterozygous SNPs for each individual was calculated as the number of heterozygous SNPs divided by the total number of genotyped SNPs for that individual. Pairwise comparisons of allele frequencies and the proportion of private alleles were calculated between groups (as defined using phylogenetic and MDS analysis), using loci that were genotyped in both groups. The pairwise allele frequencies, distribution of minor allele frequencies and proportion of shared (present in both populations) versus private (present in one group or the other) alleles are shown in Fig. 5.

Data accessibility

The following files (see supplementary materials text descriptions) are archived at Dryad (https://datadryad.org/) under the following DOIs:

1. C1AK4ACXX_5_fastq.gz;
2. Dioscorea_key_UENAK_03Oct2013.txt;
3. xml files;
4. all.mergedSNPs.vcf.gz;
5. all.mergedSNPs.GQ98.lt90pctmiss.recode.vcf;
6. all.mergedSNPs.GQ98.lt90pctmiss.tfam and all.mergedSNPs.GQ98.lt90pctmiss.tped;
7. all.mergedSNPs.GQ98.lt90pctmiss.grouped.phylip;
8. all.mergedSNPs.GQ98.lt90pctmiss.grouped.meg;
9. outweights;
10. outfile and outtree500boot;
11. outtree500boot_reroot

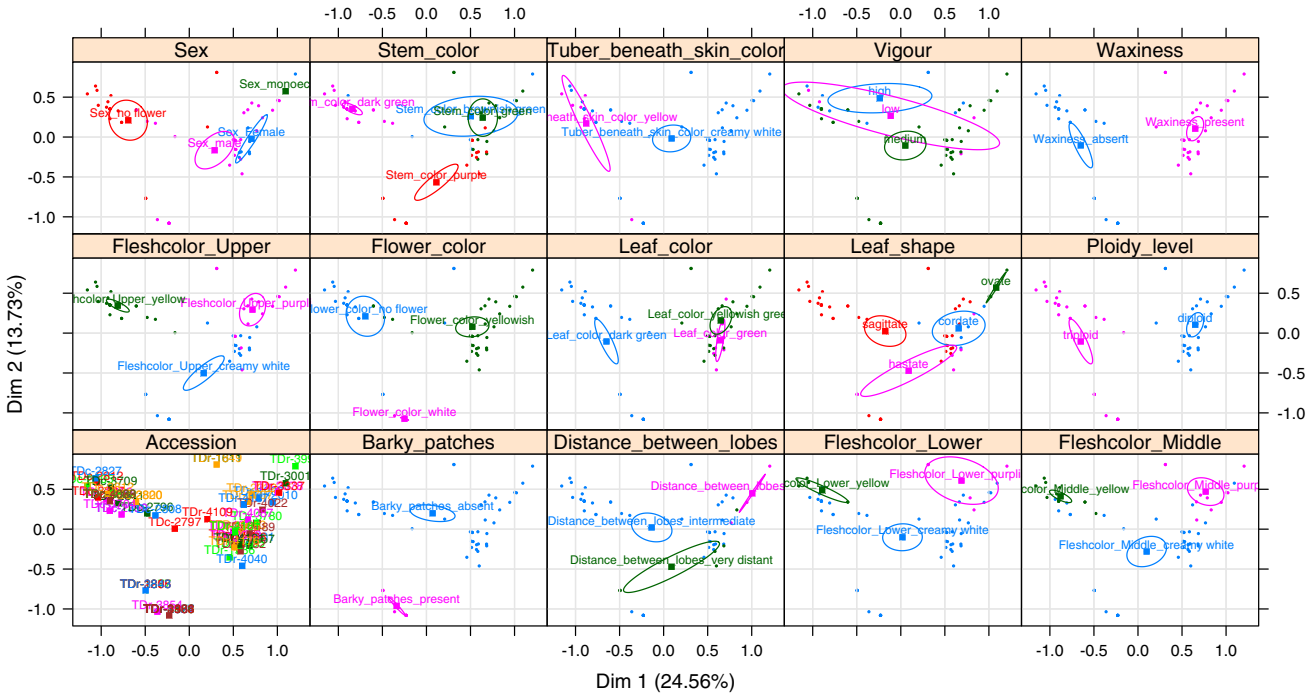

Fig. 3 Multiple correspondence analysis (MCA) performed using the plotellipses function in R, which draws confidence ellipses around the categories of all the categorical variables used. On the MCA, under accession (*bottom left*) shows a cluster of *D. cayenensis* on the *top left* and *D. rotundata* on the *top right* and *middle bottom*. For ploidy level, flesh color and some other traits the MCA shows similar pattern of variable distribution for the individuals of the two species

Results

Morphological diversity among cultivated yam species

Phenotypic descriptors have been extensively used for plant genetic resources management and conservation (Zamir 2013). Apart from tuber flesh color, none of the phenotypic descriptors used in this study could distinguish the two cultivated species from each other, although some descriptor traits correlated with ploidy level (Fig. 3; Supp Fig. 1). Morphological traits associated with increased ploidy levels in *D. rotundata* included; presence of barky patches, absence of waxiness on stem, and dark green leaf color. The yellow color of tuber flesh observed in *D. cayenensis* was absent in *D. rotundata*. *D. rotundata* was the most phenotypically diverse species in terms of flowering pattern (male, female, monoecious, and non-flowering). In *D. cayenensis*, only male or non-flowering accessions were observed. Some traits including stem color, leaf color, leaf shape, absence and presence of barky patches and waxiness, showed variation in *D. rotundata* but not in *D. cayenensis* (Fig. 3; Supp Fig. 1).

Ploidy variation across different species of guinea yams

The within-species ploidy level was constant amongst accessions of *D. cayenensis* (3x, $N = 21$), *D. praehensilis* (2x, $N = 7$), and *D. mangenotiana* (3x, $N = 5$). In contrast, both diploid (74.4 %) and triploid (25.6 %) accessions were observed for *D. rotundata* (Table 1). The coefficient of observed variation was below 5 % in all flow cytometry histograms, indicating the reliability of ploidy measurements. The different ploidy level accessions within a given species displayed differing phenotypes. For instance, triploid *D. rotundata* individuals all had distinct features, which were absent in the diploid accessions (i.e., dark green leaves, stems with barky patches and no waxiness). Moreover, all triploid (3x) individuals were either male or consistently non-flowering. All female flowering plants ($N = 8$) as well as the monoecious ($N = 1$), non-flowering ($N = 4$), and remaining male accessions ($N = 17$) were diploid (data not shown).

The genetic clustering analysis showed admixture of some individuals across different ploidy groups. Two 2x *D. rotundata* accessions were admixed with 3x *D. rotundata*, while two 2x accessions were admixed with 3x *D. cayenensis* groups (TDr 4187, TDr 3337, TDr 3990 and TDr 4087) which exhibited high heterozygosity. Ploidy variation was also associated with incorporation of alleles from wild germplasm into the *D. cayenensis–D. rotundata* complex. *D. cayenensis* (3x) harboured alleles from the wild species *D. burkilliana*, whereas 3x *D. rotundata* contained *D. togoensis* alleles, potentially indicating

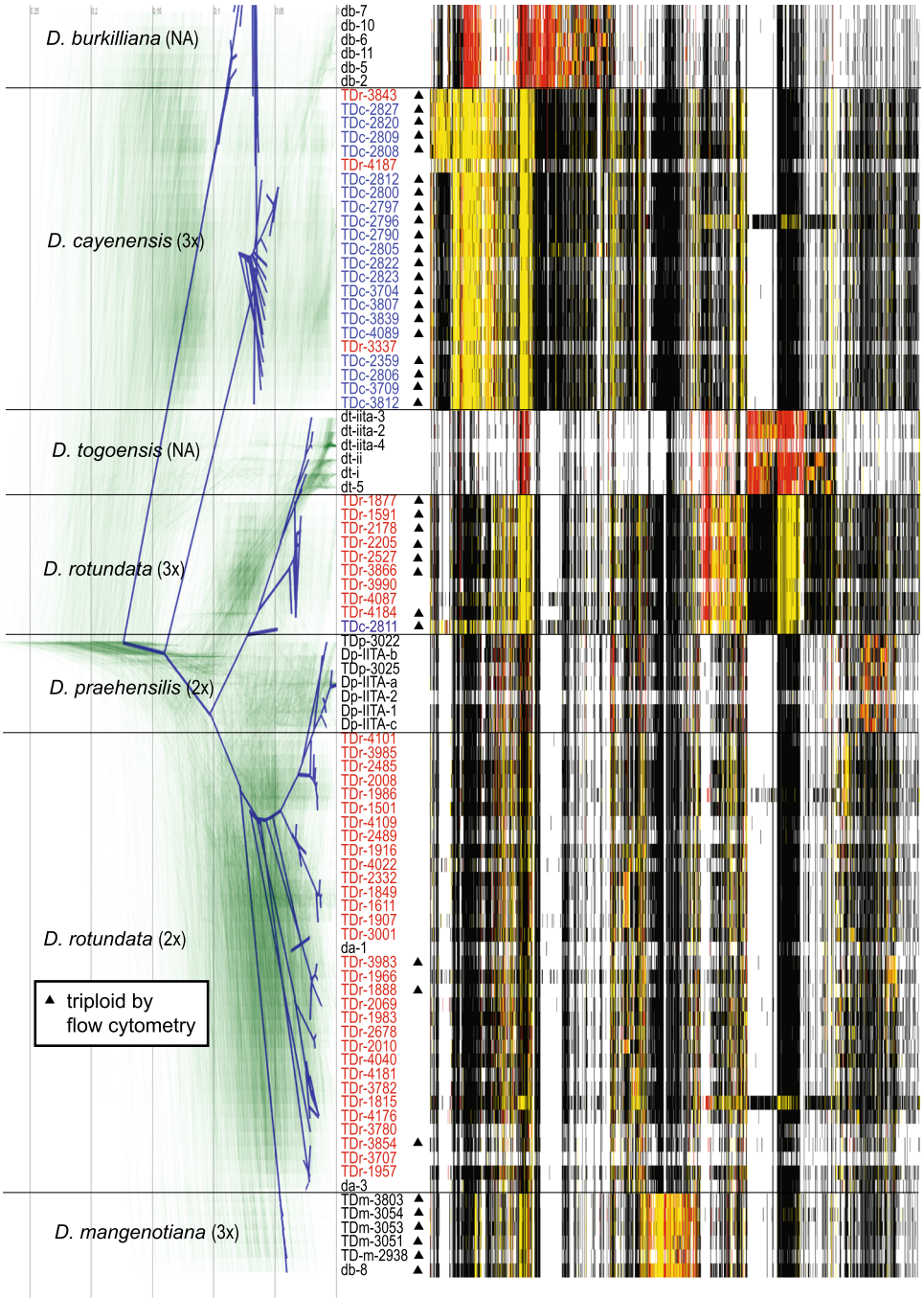

Fig. 4 Maximum parsimony analysis, genotype clustering and ploidy level. The figure shows all 25,370 maximum parsimony (MP) trees (*green*) and a single summary tree (*blue*) for 500 bootstrap MP replicates based on 2215 SNPs. Homozygous major allele are shown in *black*, homozygous minor allele are shown in *red*, heterozygous SNP are shown in *yellow*, and missing data is *left blank*. Ploidy level is shown for each individual accession where available in parenthesis after species name. Accessions from wild species are shown in *black*, *D. cayenensis* is shown in *blue*, and *D. rotundata* is shown in *red* (color figure online)

allo-polyploid origins of these 3x cultivated accessions (Figs. 4, 5). However, some 3x individuals of *D. rotundata* (TDr 3983, TDr 1888 and TDr 3854) did not have high levels of heterozygosity, indicating autopolyploidy or hybridization between closely related individuals as possible routes to polyploidy. The reduced heterozygosity in *D. burkilliana, D. togoensis* and *D. abyssinica* suggests that these are diploid.

Genetic diversity patterns and genetic structure of yams

The maximum parsimony analysis distinguished *D. burkilliana* from *D. togoensis*, but also distinguished *D. praehensilis* and *D. mangenotiana* from the cultivated *D. cayenensis* and *D. rotundata* (Fig. 4). Two *D. abyssinica* individuals appeared to be closely related to *D. rotundata* (2x) (Fig. 4). The mean group differences in substitution rate per site (Table 2)

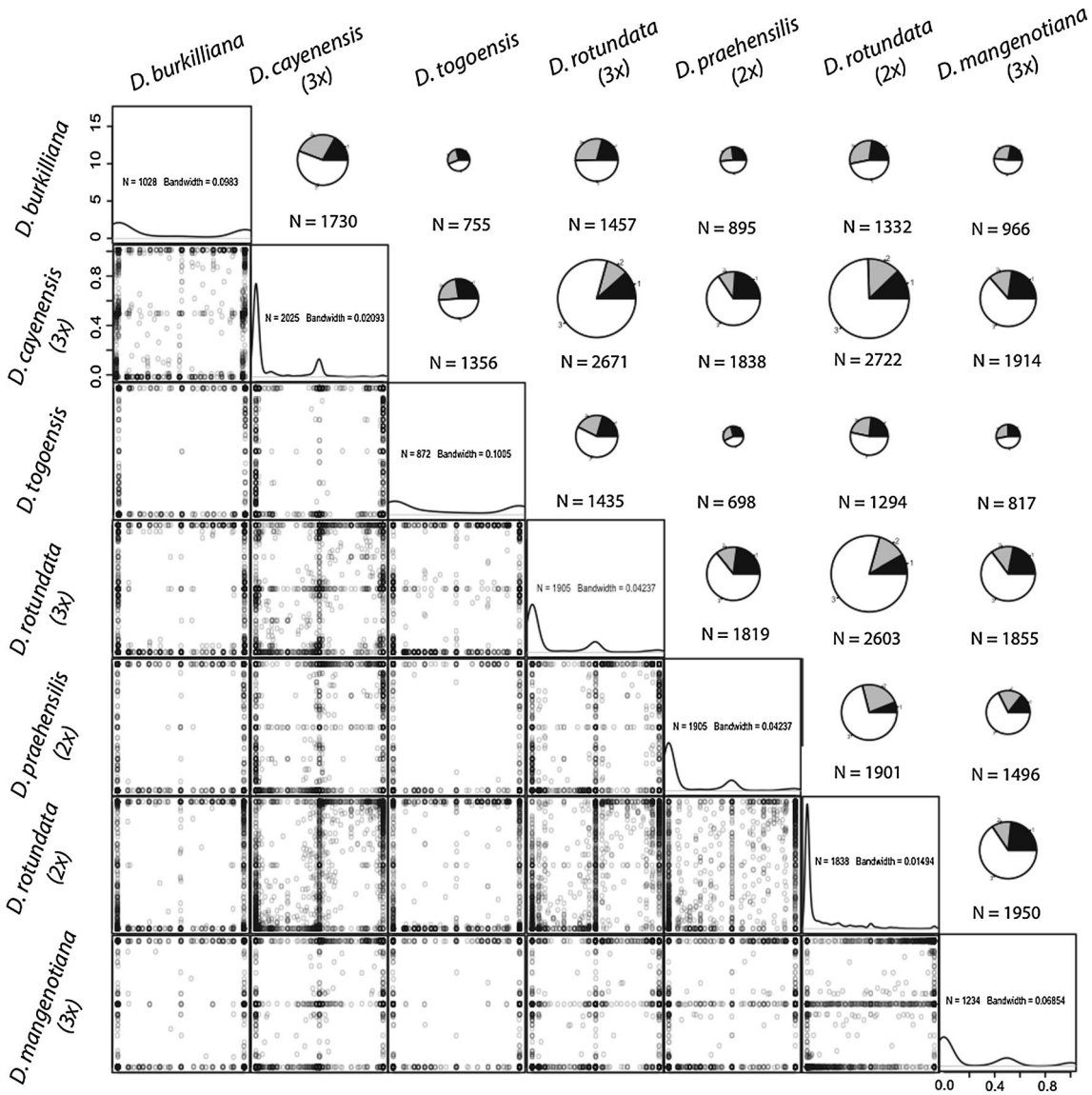

Fig. 5 Frequency and proportion of private alleles. The *lower diagonal area* contains plots of pair wise allele frequencies (major and minor) between groups. In *each box*, the points around the edges represent alleles that are fixed in one population or the other, while points in the *middle* are segregating in both. The lower diagonal area of the *figure* also shows plots of minor allele frequencies for each group (all on the same scale, 0.0–1.0)—peaks at ~50 % (0.5) can be seen in groups that are 3x. The *upper diagonal area* of the *figure* contains *pie charts* indicating the proportion of shared and private alleles (major and minor). Shared alleles represented in *white*, while private alleles specific to the *x* axis group are in *black*, and private alleles specific to the *y* axis group are in *grey*

indicated that the wild guinea yams *D. togoensis* and *D. burkilliana* are the most distant among wild populations from the cultivated species. Conversely, the analysis indicated that *D. mangenotiana*, *D. praehensilis* and *D. abyssinica* wild species are genetically closer to the cultivated species, *D. cayenensis* and *D. rotundata* (Fig. 2). The number of base substitutions per site (from averaging over all sequence pairs between groups) ranged from 0.03 between *D. abyssinica* and (2x) *D. rotundata* to 1.15 between *D. cayenensis* and *D. togoensis*.

The heterozygosity levels among individuals varied between 10 and 20 % and appeared to be correlated with ploidy levels (Fig. 6). For instance, the triploid *D. mangenotiana and D. cayenensis* had a higher proportion of heterozygous sites than the diploid *D. rotundata* and *D. praehensilis*. The wild yam species formed some discrete genetic groupings (Fig. 2), with *D. burkilliana* and *D. togoensis* being quite distinct from the other species. All other accessions clustered into three groups predominantly composed of the cultivated *D. rotundata* and *D. cayenensis* diploids and triploids (Fig. 2). However, accessions from the wild species *D. praehensilis*, *D. mangenotiana* and *D. abyssinica* clustered together with the cultivated species.

Table 2 Estimates of evolutionary divergence over sequence pairs between groups

	abyssinica	*burkilliana*	*mangenotiana*	*praehensilis*	*togoensis*	*cayenensis*	*rotundata_2x*	*rotundata_3x*
abyssinica		0.1058505	0.0098201	0.0139287	0.1537379	0.0090677	0.0032199	0.0054628
burkilliana	0.8032981		0.0918105	0.0986323	0.1265432	0.0597261	0.0899553	0.1023735
mangenotiana	0.0582308	0.7092618		0.0165598	0.1549088	0.0182714	0.0091955	0.0140782
praehensilis	0.1165054	0.8175970	0.1354147		0.1510850	0.0138823	0.0102911	0.0131464
togoensis	0.9549081	0.8918185	1.0429453	1.0149002		0.1929068	0.1369335	0.0689549
cayenensis	0.0524431	0.5150288	0.1452489	0.1193613	1.1551302		0.0059603	0.0099503
rotundata_2x	0.0304090	0.8032775	0.0792767	0.1174252	0.9331154	0.0584111		0.0039185
rotundata_3x	0.0425668	0.8452944	0.1233333	0.1244797	0.5777506	0.0900501	0.0449827	

Lower left diagonal: average number of base substitutions per site over all sequence pairs between groups. Upper right diagonal: standard error estimate(s) are shown above the diagonal

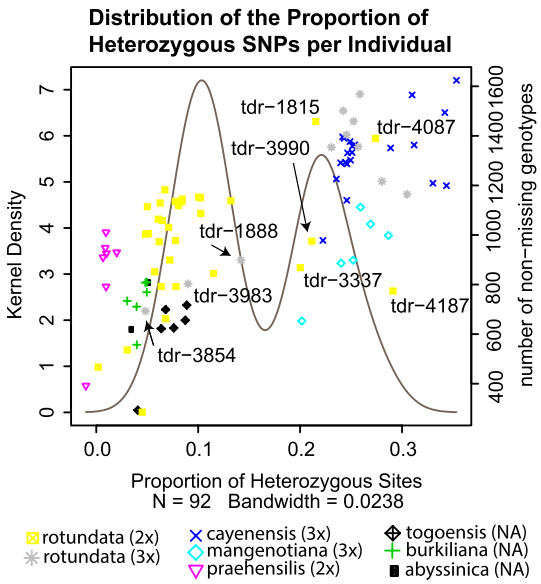

Fig. 6 Kernel density (probability density function) of the proportion of heterozygous SNPs per individual across seven guinea yam species. Ploidy levels are also indicated

Discussion

Identification of novel guinea yam SNPs through application of GBS

GBS is increasingly used for genetic diversity analyses, gene identification, and plant breeding. GBS has been applied to wheat genomic selection (Poland et al. 2012a), analysis of switchgrass genomic diversity (Lu et al. 2013), development of genetic maps in barley and wheat (Poland et al. 2012b), and genome wide association studies in sorghum (Morris et al. 2013). Here we demonstrate that GBS is an effective tool for analysis of guinea yam genomic diversity, regardless of the complexity of guinea yams in terms of ploidy level, genome size, and the current lack of a reference genome.

Recent origins of cultivated yams from wild ancestors such as *D. burkilliana*

The low genetic divergence between the two cultivated species, *D. rotundata* and *D. cayenensis* (Table 2), confirms previous studies (using RFLP analysis) which suggested that these two species display a recent evolutionary divergence (Terauchi et al. 1992). The clear separation of *D. togoensis* and *D. burkilliana* illustrates the isolation of these species from the *rotundata–cayenensis* complex. The relatively lower divergence (Table 2) and higher allele sharing (Fig. 4) between *D. cayenensis* and *D. burkilliana* substantiates earlier suggestions (Onyilagha and Lowe 1986; Ramser et al. 1997; Terauchi et al. 1992) that *D. burkilliana* could be the possible ancestor of *D. cayenensis*. However, *D. togoensis* seems to contribute more to *D. rotundata* (based on allele sharing) than to *D. cayenensis*, in contrast to previous reports (Ramser et al. 1997; Terauchi et al. 1992). The minimal differentiation and closer similarity of *D. mangenotiana*, *D. praehensilis* and *D. abyssinica* (Fig. 2; Table 2) with the *rotundata-cayenensis* complex indicates that these wild relatives are either of recent divergence or variants of the cultivated species.

Population genetic structure of the cultivated guinea yams and its wild relatives likely reflects ongoing domestication practices or past hybridization events

The wild relatives of yam display distinct clustering based on multi-dimensional scaling, maximum parsimony and genotype clustering (Figs. 4, 2). However, some of the wild relatives showed some genetic admixture with cultivated forms. The close genetic relationship between the wild and cultivated species could also be due to the difficulty to phenotypically differentiate the cultivated species from the wild species, or due to gene flow occurring via interspecific hybridization between wild and cultivated species (Cornet et al. 2010; Scarcelli et al. 2006a, b).

Our study also shows evidence of admixture in *D. abyssinica* (Fig. 4). The spontaneous formation of hybrids between wild and cultivated yams demonstrated by Scarcelli et al. (2006a) suggests a mechanism for naturally occurring genetic admixture between cultivated and wild relatives. In contrast, Dansi et al. (1999) found no evidence for the deliberate use of *D. togoensis* plants for domestication purposes. There is also no report of farmers harvesting *D. burkilliana* for food or for other domestication purposes, although farmers do recognize both of these species as wild (Dansi et al. (1999). Ethnobotanical evidence suggests that gene flow between these two wild species and cultivated yams is minimal. This is supported by our data, which showed little genetic contribution of these two wild species to the cultivated gene pool (*rotundata–cayenensis* complex). The increased heterozygosity levels we found in some 2x accessions for *D. rotundata* also supports a role for admixture arising from interspecific hybridization.

Breeding (crossing) experiments conducted at IITA have confirmed the sexual compatibility within cultivated, and between cultivated yams and their wild relatives. Interspecific crossing studies were conducted with the objective to transfer traits from wild relatives to cultivated lines. Cultivated species (e.g. *D. rotundata* × *D. cayenensis)* and wild and cultivated species (e.g. *D. rotundata* × *D. praehensilis* and *D. rotundata* × *D. togenesis*) have been successfully crossed (Akoroda 1985); Robert Asiedu, personal communication).

Morphological descriptors lack resolving power to differentiate the two cultivated yam species

The taxonomy of *D. rotundata* and *D. cayenensis* has been under investigation and scientific debate for decades. Some taxonomists have considered *D. rotundata* as subspecies of *D. cayenensis,* indicated as *D. cayenensis subsp. rotunda* (Poiret) J. Miège 1968 (White Guinea Yam) whereas Terauchi et al. (1992) suggested that 'yellow yam', *D. cayenensis* should be treated as a variety of *D. rotundata*, denoted as *D. rotundata var.* × '*cayenen-sis*' *(*on the basis of its nuclear ribosomal DNA characteristics). On the other hand, Hamon and Toure (1990) observed several intermediate forms, and proposed to treat the two species as the *D. cayenensis–rotundata* species complex.

In this study, we have observed yellow tuber flesh color in some parts of the tuber in all *D. cayenensis* accessions investigated (Fig. 4). However, as a classifier, the yellow tuber flesh color is ambiguous in some accessions of *D. cayenensis* even though it is the most commonly used approach for classifying the two species as either yellow or white yams. Illustrating the challenges of using morphological descriptors, none of the morphological descriptors we used were distinct for the two yam species highlighting the difficulty to distinguish the two species using such criteria. However, our analysis determined that some of the morphological traits are correlated with ploidy level. For instance, the presence of barky patches, absence of waxiness, and dark green leaf color are closely related with 3x *D. rotundata.*

Ploidy variation in guinea yams due to auto- and allo-polyploidy

The pattern of allele sharing where *D. cayenensis* harboured *D. burkilliana* alleles, 3x *D. rotundata* harboured *D. togoensis* alleles and a few 3x *D. rotundata* showed reduced heterozygosity, suggest that the polyploidization process in guinea yams likely involves both allo-polyploidy and auto-polyploidy. Moreover, the increased heterozygosity in some 2x *D. rotundata* accessions highlights the presence of gene flow between closely related species. Additionally, increased ploidy levels and heterozygosity in *D. cayenensis* and allele sharing between the two cultivated species indicate that *D. cayenensis* arose from *D. rotundata* but not vice versa. Our results also confirm the earlier suggestion of Terauchi et al. (1992) to consider *D. cayenensis* as a subspecies of *D. rotundata*.

Guinea yam GBS data will be most powerful when combined with reference genome

Despite the lack of a reference genome, the UNEAK pipeline was successfully used to call large number of SNPs in switchgrass (Lu et al. 2013), which were further validated using maize GBS data. While we have utilised the GBS data in the absence of a reference genome for yam, we recognize that GBS is most powerful when the reference genome is available. Access to a reference genome for yam would help for identifying more SNPs and avoiding potential bias associated with the conservative SNP calling employed in the UNEAK pipeline. The GBS data generated in this study (and made publicly available) will be compatible with the yam reference genome when the genome sequence is released, and will allow further assessment of molecular diversity in yam.

Specifically, the limitation of identifying bi-allelic SNPs that differ at only one base pair within a 64 base pair tag may lead to biases when estimating true rates of divergence within or among species due to mis-identified or unobserved loci, especially when divergence rates are high. This problem may also be exacerbated by low sample sizes for some species. With a reference genome, these biases can be significantly reduced. The GBS raw sequence data generated in this study will be reanalyzed in the future using a reference-sequence based pipeline for calling SNPs once the genome sequence of *Dioscorea* becomes available (Tamiru et al. 2013).

Implications for guinea yam conservation and improvement programs

We advocate the wider use of GBS (even in species lacking a reference genome), as it can help generating genotypic information across the whole population of interest (including germplasm collections) at a much lower cost per data point. Similarly, GBS could be used for further understanding of genetic relationship studies of other species within the genus *Dioscorea*. GBS is cost-effective and has major potential for characterization of the yam genebank collection maintained at IITA (and other yam germplasm collections), as it can assess the extent and distribution of genetic diversity in the collections. Such knowledge is necessary for improved management of the genebank, either through identifying duplicates or guiding the need for further germplasm collection. The close genetic similarity of some wild yams with the cultivated forms and sexual compatibility between species, provides an opportunity for yam improvement through incorporation of genes and traits from wild relatives. The use of wild relatives in yam breeding programs can allow the tapping of important traits present in the wild genetic pool and that were not yet captured in domesticated germplasm. Variation in ploidy within and between species is a challenge but also an opportunity for managing both intraspecific and interspecific hybridization in breeding programs. Overall, the use of GBS combined with a better understanding of ploidy relationships among species is essential for improving understanding of genetic relationships between wild and cultivated forms of guinea yams, which is critical for understanding the evolution, domestication and ongoing use of guinea yams as an important staple food crop.

Acknowledgments The CGIAR Research Program on Roots, Tubers and Bananas funded this research. Prof. Alexander Dansi provided materials from Benin. Prof. Michael Abberton and Dr. Marc Deletre reviewed the manuscript. GG acknowledges support from the Netherlands Ministry of Foreign Affairs, and PhD fee waiver from the National University of Ireland Galway. CS acknowledges support from Science Foundation Ireland (SFI).

Conflict of interest The authors declare that they have no conflict of interest.

References

Aidoo R, Nimoh F, Bakang J-EA, Ohene-Yankyera K, Fialor SC, Abaidoo RC (2011) Economics of small-scale seed yam production in Ghana: implications for commercialisation. J Sustain Dev Africa 13:65–78

Akoroda MO (1985) Pollination management for controlled hybridization of white yam. Sci Hortic 25:201–209

Amusa N, Adegbite A, Mohammed S, Baiyewu R (2003) Yam diseases and its management in Nigeria. Afr J Biotechnol 2:497–502

Arnau G, Nemorin A, Maledon E, Abraham K (2009) Revision of ploidy status of *Dioscorea alata* L. (Dioscoreaceae) by cytogenetic and microsatellite segregation analysis. Theor Appl Genet 118:1239–1249

Babil PK, Irie K, Shiwachi H, Tun Y, Toyohara H, Fujimaki H (2010) Ploidy variation and their effects on Leaf and stoma traits of water yam (*Dioscorea alata* L.) collected in Myanmar. Trop Agric Dev 54:132–139

Bahuchet S, McKey D, de Garine I (1991) Wild yams revisited: is independence from agriculture possible for rain forest hunter-gatherers? Hum Ecol 19:213–243

Baimey H, Coyne D, Labuschagne N (2006) Effect of fertilizer application on yam nematode (*Scutellonema bradys*) multiplication and consequent damage to yam (*Dioscorea* spp.) under field and storage conditions in Benin. Int J Pest Manag 52:63–70

Bouckaert RR (2010) DensiTree: making sense of sets of phylogenetic trees. Bioinformatics 26:1372–1373

Bousalem M, Arnau G, Hochu I, Arnolin R, Viader V, Santoni S, David J (2006) Microsatellite segregation analysis and cytogenetic evidence for tetrasomic inheritance in the American yam *Dioscorea trifida* and a new basic chromosome number in the Dioscoreae. Theor Appl Genet 113:439–451

Burkill IH (1960) The organography and the evolution of Dioscoreaceae, the family of the yams. J Linn Soc Lond Bot 56:319–412

Cornet D, Deu M, Baco M, Agbangla A, Duval M, Noyer J (2010) Impact of farmer selection on yam genetic diversity. Conserv Genet 11:2255–2265

Coursey DG (1967) Yams. An account of the nature, origins, cultivation and utilisation of the useful members of the Dioscoreaceae. Longmans, Green and Co. Ltd, London

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Group GPA (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158

Dansi A, Mignouna H, Zoundjihekpon J, Sangare A, Asiedu R, Quin F (1999) Morphological diversity, cultivar groups and possible descent in the cultivated yams (*Dioscorea cayenensis/D. rotundata*) complex in Benin Republic. Genet Resour Crop Evol 46:371–388

Dansi A, Mignouna HD, 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

Deschamps S, Llaca V, May GD (2012) Genotyping-by-sequencing in plants. Biology 1:460–483

Development Core Team R (2010) R: a language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria

Dumet D, Ogunsola D (2008) Regeneration guidelines: yams, crop specific regeneration guidelines. In: Dulloo ME, Thormann I, Jorge MA, Hanson J (eds) CGIAR System-wide Genetic Resource Programme, Rome

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379

Emshwiller E (2002) Ploidy levels among species in the '*Oxalis tuberosa* Alliance' as inferred by flow cytometry. Ann Bot 89:741–753

FAOSTAT (2013) Statistical data base. FAO, Rome

Gamiette F, Bakry F, Ano G (1999) Ploidy determination of some yam species (*Dioscorea* spp.) by flow cytometry and conventional chromosomes counting. Genet Resour Crop Evol 46:19–27

Govaerts R, Wilkin P, Saunders RMK (2007) World checklist of Dioscoreales. Yams and their allies. The Board of Trustees of the Royal Botanic Gardens, Kew, London, pp 1–65

Hamon P, Toure B (1990) Characterization of traditional yam varieties belonging to the *Dioscorea cayenensis–rotundata* complex by their isozymic patterns. Euphytica 46:101–107

Hamon P, Brizard J-P, Zoundjihékpon J, Duperray C, Borgel A (1992) Etude des index d' ADN de huit ignames (*Dioscorea* sp.) par cytométrie en flux. Can J Bot 70:996–1000

IPGRI/IITA (1997) Descriptors for yam (*Dioscorea* spp.). International Institute of Tropical Agriculture, Ibadan, Nigeria/International Plant Genetic Resources Institute, Rome, Italy

Junier T, Zdobnov EM (2010) The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. Bioinformatics 26:1669–1670

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinf 9:559

Lê S, Josse J, Husson F (2008) FactoMineR: an R package for multivariate analysis. J Stat Softw 25:1–18

Lebot V (2009) Tropical root and tuber crops: cassava, sweet potato, yams and aroids. CABI Publishers, Wallingford

Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Buckler ES, Costich DE (2013) Switchgrass genomic diversity, ploidy and evolution: novel insights from a network-based SNP discovery protocol. PLoS Genet 9:e1003215

Mengesha W, Demissew S, Fay M, Smith R, Nordal I, Wilkin P (2013) Genetic diversity and population structure of Guinea yams and their wild relatives in South and South West Ethiopia as revealed by microsatellite markers. Genet Resour Crop Evol 60:529–541

Mignouna HD, Dansi A, Zok S (2002) Morphological and isozymic diversity of the cultivated yams (*Dioscorea cayenensis*/*Dioscorea rotundata* complex) of Cameroon. Genet Resour Crop Evol 49:21–29

Mignouna HD, Abang MM, Fagbemi SA (2003) A comparative assessment of molecular marker assays (AFLP, RAPD and SSR) for white yam (*Dioscorea rotundata*) germplasm characterization. Ann Appl Biol 142:269–276

Mignouna HD, Abang MM, Asiedu R (2007) Yams. In: Kole C (ed) Genome mapping and molecular breeding: pulses, sugar and tuber crops. Springer, Berlin, Heidelberg, pp 271–296

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE, Harriman J, Glaubitz JC, Buckler ES, Kresovich S (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc Natl Acad Sci USA 110:453–458

Nemorin A, Abraham K, David J, Arnau G (2012) Inheritance pattern of tetraploid *Dioscorea alata* and evidence of double reduction using microsatellite marker segregation analysis. Mol Breed 30:1657–1667

Obidiegwu J, Loureiro J, Ene-Obong E, Rodriguez E, Kolesnikova-Allen M, Santos C, Mouneke C, Asiedu R (2009) Ploidy level studies on the *Dioscorea cayenensis*/*Dioscore rotundata* complex core set. Euphytica 169:319–326

Onyilagha JC, Lowe J (1986) Studies on the relationship of *Dioscorea cayenensis* and *Dioscorea rotundata* cultivars. Euphytica 35:733–739

Poland JA, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. Plant Gen 5:92–102

Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sánchez-Villeda H, Sorrells M, Jannink J-L (2012a) Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Gen 5:103–113

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012b) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7:e32253

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575

R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria. http://www.R-project.org/

Ramser J, Weising K, Terauchi R, Kahl G, Lopez-Peralta C, Terhalle W (1997) Molecular marker based taxonomy and phylogeny of Guinea yam (*Dioscorea rotundata–D. cayenensis*). Genome/National Res Council Canada = Genome/Conseil national de recherches Canada 40:903–915

Sato H (2001) The potential of edible wild yams and yam-like plants as a staple food resource in the African tropical rain forest. Afr Study Monogr Suppl Issue 26:123–134

Scarcelli N, Dainou O, Agbangla C, Tostain S, Pham JL (2005) Segregation patterns of isozyme loci and microsatellite markers show the diploidy of African yam *Dioscorea rotundata* ($2n = 40$). Theor Appl Genet 111:226–232

Scarcelli N, Tostain S, Mariac C, Agbangla C, Da O, Berthaud J, Pham JL (2006a) Genetic nature of yams (*Dioscorea* sp.) domesticated by farmers in Benin (West Africa). Genet Resour Crop Evol 53:121–130

Scarcelli N, Tostain S, Vigouroux Y, Agbangla C, DaÏNou O, Pham JL (2006b) Farmers' use of wild relative and sexual reproduction in a vegetatively propagated crop. The case of yam in Benin. Mol Ecol 15:2421–2431

Scarcelli N, Couderc M, Baco M, Egah J, Vigouroux Y (2013) Clonal diversity and estimation of relative clone age: application to agrobiodiversity of yam (*Dioscorea rotundata*). BMC Plant Biol 13:178

Spillane C, Gepts P (2001) Evolutionary and genetic perspectives on the dynamics of crop genepools. In: Cooper D, Spillane C, Hodgkin T (eds) Broadening the genetic base of crop production. CABI, Wallingford, pp 25–70

Spindel J, Wright M, Chen C, Cobb J, Gage J, Harrington S, Lorieux M, Ahmadi N, McCouch S (2013) Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theoret Appl Genet 126:2699–2716

Tamiru M, Natsume S, Takagi H, Babil PK, Yamanaka S, Lopez-Montes A, Gedil M, Bhattacharjee R, Takagi H, Asiedu R, Terauchi R (2013) Whole genome sequencing of Guinea yam (*Dioscorea rotundata*). In: First global conference on yam. International Institute of Tropical Agriculture (IITA), Accra, Ghana

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Terauchi R, Chikaleke VA, Thottappilly G, Hahn SK (1992) Origin and phylogeny of Guinea yams as revealed by RFLP analysis of chloroplast DNA and nuclear ribosomal DNA. Theoret Appl Genet 83:743–751

Tortoe C, Johnson P-T, Abbey L, Baidoo E, Anang D, Acquaah SG, Saka E (2012) Sensory properties of pre-treated blast-chilled (*Dioscorea rotundata*) as a convenience food product. Afr J Food Sci Technol 3:59–65

Zamir D (2013) Where have all the crop phenotypes gone? PLoS Biol 11:e1001595

Zannou A, Ahanchede A, Struik P, Richard P, Zoundjihekpon J, Tossou R, Vodouhe S (2004) Yam and cowpea diversity management by farmers in the Guinea–Sudan transition zone of Benin. NJAS (Wageningen J Life Sci) 52:393–420

Zannou A, Richards P, Struik PC (2006) Knowledge on yam variety development: insights from farmers' and researchers' practices. Knowl Manag Dev J 2:30–39