



**HAL**  
open science

## **Development of a cost-effective single nucleotide polymorphism genotyping array for management of greater yam germplasm collections**

Fabien Cormier, Pierre Mournet, Sandrine Causse, Gemma Arnau, Erick Maledon, Rose-Marie Gomez, Claude Pavis, Hana Chair

### **► To cite this version:**

Fabien Cormier, Pierre Mournet, Sandrine Causse, Gemma Arnau, Erick Maledon, et al.. Development of a cost-effective single nucleotide polymorphism genotyping array for management of greater yam germplasm collections. *Ecology and Evolution*, 2019, 9 (10), pp.5617 - 5636. <10.1002/ece3.5141>. <hal-02625995>

**HAL Id: hal-02625995**

**<https://hal.inrae.fr/hal-02625995v1>**

Submitted on 26 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.


L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

## ORIGINAL RESEARCH

# Development of a cost-effective single nucleotide polymorphism genotyping array for management of greater yam germplasm collections

Fabien Cormier<sup>1,2</sup>  | Pierre Mournet<sup>2,3</sup> | Sandrine Causse<sup>2,3</sup> | Gemma Arnau<sup>1,2</sup> | Erick Maledon<sup>1,2</sup> | Rose-Marie Gomez<sup>4</sup> | Claudie Pavis<sup>4</sup> | Hâna Chair<sup>2,3</sup>

<sup>1</sup>CIRAD, UMR AGAP, Petit-Bourg, France

<sup>2</sup>CIRAD, INRA, Univ Montpellier, Montpellier SupAgro, Montpellier, France

<sup>3</sup>CIRAD, UMR AGAP, Montpellier, France

<sup>4</sup>INRA, UR ASTRO Agrosytèmes Tropicaux, Petit-Bourg, France

## Correspondence

Fabien Cormier, CIRAD, UMR AGAP, Petit-Bourg, France.

Email: fabien.cormier@cirad.fr

## Funding information

European Union and Guadeloupe Region

## Abstract

Using genome-wide single nucleotide polymorphism (SNP) discovery in greater yam (*Dioscorea alata* L.), 4,593 good quality SNPs were identified in 40 accessions. One hundred ninety six of these SNPs were selected to represent the overall dataset and used to design a competitive allele specific PCR array (KASPar). This array was validated on 141 accessions from the Tropical Plants Biological Resources Centre (CRB-PT) and CIRAD collections that encompass worldwide *D. alata* diversity. Overall, 129 SNPs were successfully converted as cost-effective genotyping tools. The results showed that the ploidy levels of accessions could be accurately estimated using this array. The rate of redundant accessions within the collections was high in agreement with the low genetic diversity of *D. alata* and its diversification by somatic clone selection. The overall diversity resulting from these 129 polymorphic SNPs was consistent with the findings of previously published studies. This KASPar array will be useful in collection management, ploidy level inference, while complementing accurate agro-morphological descriptions.

## KEYWORDS

*Dioscorea alata* L., ex situ collection, genotyping, KASPar, ploidy, yam

## 1 | INTRODUCTION

Greater yam (*Dioscorea alata* L.) is one of the major cultivated yam species (*Dioscorea* spp.) and the most widely spread among tropical and subtropical regions. The high importance of *D. alata* for food security has prompted the establishment of several international and national ex situ collections. Due to the limited shelf-life of stored tuber, yam genetic resources are conserved in vitro or/and in the field. All of these repeated manipulations are time-consuming and may affect long-term conservation. Quality control of genotype purity and general collection management is mainly based on morphological

descriptors (IPGRI/IITA, 1997; Mahalakshmi et al., 2007). However, these descriptors are not reliable enough to rationalize ex situ *D. alata* collection. Indeed, several studies have revealed that morphological variations are not necessarily linked to geographic origin or genetic lineage (Arnau et al., 2017; Lebot, Trilles, Noyer, & Modesto, 1998; Vandenbroucke et al., 2016). Complementary characterization tools are thus required for the conservation and dynamic management of ex situ collections related to germplasm exchange, the development of core collection or identification of future parents for breeding programs. *D. alata* is also a polyploid species with ploidy levels of  $2n = 2x$ ,  $3x$ , or  $4x$  and a basic chromosome number of  $x = 20$  (Arnau, Némorin,

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

Maledon, & Abraham, 2009). Ploidy levels detection is consequently a prerequisite for the identification of possible parents as crosses between the different ploidy levels can fail (Nemorin et al., 2013).

Molecular markers have been used to characterize *D. alata* diversity: random amplified polymorphic DNA (RAPD; Asemota, Ramser, Lopez-Peralta, Weising, & Kahl, 1996), isoenzymes (Lebot et al., 1998), amplified fragment length polymorphism (AFLP; Malapa, Arnau, Noyer, & Lebot, 2005), simple sequence repeats (SSRs; Siqueira, Marconi, Bonatelli, Zucchi, & Veasey, 2011; Sartie, Asiedu, & Franco, 2012; Otoo, Anokye, Asare, & Telleh, 2015; Chair et al., 2016; Arnau et al., 2017), plastid sequences (Chair et al., 2016), and Diversity Arrays Technology (DArT; Vandenbroucke et al., 2016). These studies generated essential information on the diversity and representativity of the germplasm collections. However, these tools were not tailored for routine collection management. They were found to be either poorly discriminating within *D. alata* species or they were complex and not cost-effective to use. Besides the development of high-throughput methods for genome-wide variant detection, such as genotyping-by-sequencing (Davey et al., 2011) paired with cost-effective SNP assay (Broccanello et al., 2018) as KASPar can lead to the development of appropriate markers for collection management. This approach has been successfully implemented in maize (Semagn et al., 2012), chickpea (Hiremath et al., 2012), *Citrus* (Garcia-Lor, Ancillo, Navarro, & Ollitrault, 2013), pigeon pea (Saxena et al., 2014), and *Brassica rapa* (Su et al., 2018). Regarding the recent release of yam (*Dioscorea* spp.) genomic resources (Saski, Bhattacharjee, Scheffler, & Asiedu, 2015; Tamiru et al., 2017), the design of such markers for *D. alata* collection management would be worthwhile. Indeed, once developed they do not require any specific bioinformatics or wet chemistry skills. The results contain few erroneous and missing data and can be easily analyzed and interpreted.

The main objectives of this study were (a) to identify genome-wide polymorphic SNP markers, (b) to develop a cost-effective SNP genotyping array using KASPar technology and (c) to test its use as a tool in managing yam ex situ collections.

## 2 | MATERIALS AND METHODS

### 2.1 | Materials

Based on a previous microsatellite markers study (Arnau et al., 2017), a set of 48 accessions representing worldwide *D. alata* diversity was selected and genotyped to identify polymorphic SNPs and design KASPar markers. Then, for the purpose of validating these markers, 141 landraces from the Tropical Plants Biological Resources Centre (CRB-PT) and CIRAD ex situ collections maintained in the West French Indies (Guadeloupe) were used.

### 2.2 | Genotyping-by-sequencing (GBS) and SNP discovery

SNP discovery was based on genotyping-by-sequencing (GBS). First, DNA extractions were performed with dried leaves from the

48 accessions as described by Risterucci et al. (2009). The genomic DNA quality was checked using agarose gel electrophoresis, and the quantity was estimated using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, USA). For GBS, a genomic library was prepared using the PstI-MseI restriction enzymes (New England Biolabs, Hitchin, UK) with a DNA normalized quantity of 200 ng per sample. The procedures published by Elshire et al. (2011) were adapted as described in Cormier et al. (2019).

Digestion and ligation reactions were conducted in the same plate. Digestion was conducted at 37°C for 2 hr and then 65°C for 20 min to inactivate the enzymes. The ligation reaction was achieved using T4 DNA ligase enzyme (New England Biolabs, Hitchin, UK) at 22°C for 1 hr, and the ligase was then inactivated, prior to sample pooling, by heating at 65°C for 20 min. Pooled samples were PCR-amplified in a single tube. Single-end sequencing was performed on a paired-end lane of an Illumina HiSeq3000 (at the GeT-PlaGe platform, Toulouse, France). The Tassel 5.2 pipeline (Glaubitz et al., 2014) was used for SNP and indel calling. Sequence tags were aligned to *D. alata* contigs (<http://www.ebi.ac.uk/ena/data/view/PRJEB10904>) using Bowtie2 v2.2.6 (Langmead & Salzberg, 2012). Accessions with more than 70% missing data were removed. Vcf filtering was performed using Vcftools 0.1.14 (Danecek et al., 2011; option: --minDP 8, --maf 0.1, --max-missing 0.60, --max-alleles 2, --thin64).

### 2.3 | KASPar genotyping and allele calling

Polymorphic SNP flanking sequences (60 bp upstream and 60 bp downstream around the variant position) were selected using SNIPlay3 (Dereeper et al., 2011). In order to assess their putative physical positions, these sequences were then blasted to the *D. rotundata* reference genome (TDr96\_F1 Pseudo\_Chromosome: BDMI01000001–BDMI01000021; Tamiru et al., 2017). The physical position of each SNP was defined using their flanking sequences best hit using a BLAST E-value threshold of  $1e^{-30}$  (Basic Local Alignment Search Tool). Finally, 192 SNPs were selected by forming 192 k-means cluster based on their relative physical distance (Euclidean distance) and selecting the SNP nearest to the centroid of each cluster using R 3.4.0 (R core team, 2017).

The 192 SNPs were converted into a KASPar assay at LGC genomics where the primer design and wet chemistry was conducted (Middlesex, UK) on a validation panel of 141 landraces from the CRB-PT and CIRAD ex situ collections. From raw fluorescence data, allele calling was performed using LGC Kuster Caller software by defining fluorescence clusters. Some accessions with known ploidy level were used as reference to identify fluorescence clusters and assess allelic dosage.

### 2.4 | Diversity analysis

To identify duplicate accessions and compare accessions with different ploidy levels, a matrix of dissimilarity between each accession pair was computed as the percentage of shared alleles based on the allele presence/absence.

Then, to refine the kinship assessment, similarities between accessions with the same ploidy level were computed in the same way but using the allelic dosage. For diploid accessions, genotypes were coded as 0, 1, and 2 where the number represents the number of nonreference allele. Heterozygous genotypes assessed as polyploid during allele calling were converted to 1. Moreover, for triploid accessions, genotypes were coded as 0, 1, 2, and 3 with allelic dosage score as 1:1 during allele call converted to 1.5. For tetraploid accessions, genotypes were thus coded as 0, 1, 2, 3, or 4 and no correction was needed.

Diversity analysis was conducted in two steps. During the first step, groups of duplicate accessions (redundancy groups) were defined by grouping accessions having up to one allele mismatch. Then, in the second step, the diversity analysis focused on the similarity between those groups. Clustering based on allele frequencies within redundancy groups followed by a bootstrap approach (pvclust R package, ward.D2, 10,000 boots, AU threshold = 0.95; Suzuki & Shimodaira, 2006) was used to identify gene pools. A diversity network between redundancy groups was also drawn using significant kinship detected through genotype permutations (1,000), with a significance threshold of 0.05.

### 3 | RESULTS

#### 3.1 | KASPar assay development and validation

Genotyping-by-sequencing (GBS) produced more than 344 million reads resulting in 521,918 sequence tags out of which 207,810 (39.82%) aligned exactly once on *D. alata* contigs. The remaining reads aligned at multiple locations (25.18%) or did not align to any contig (35%). From these sequence tags, SNP calling produced a raw

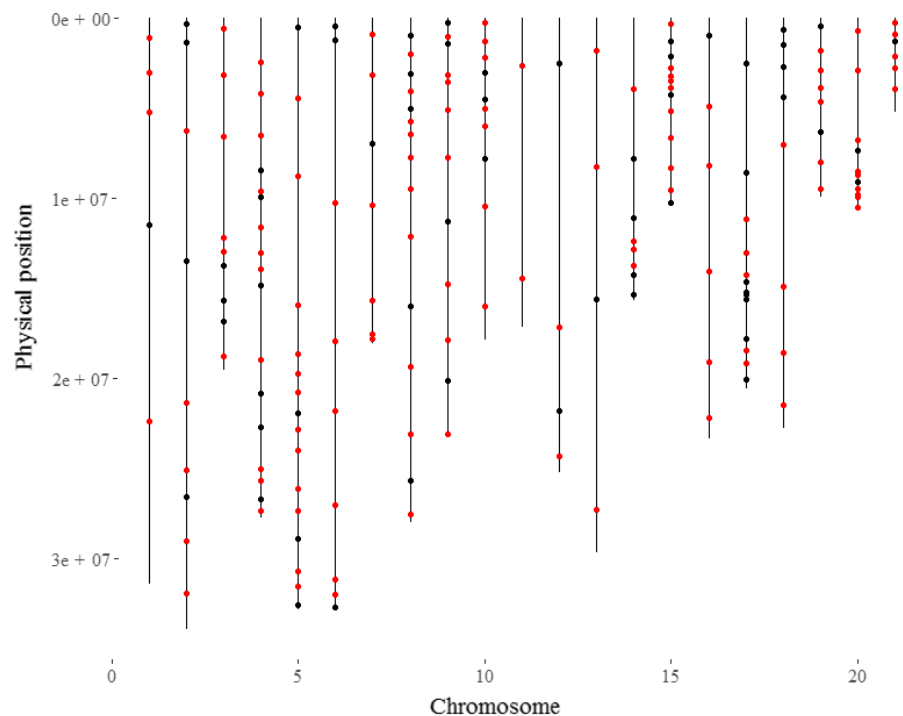
vcf file of 158,695 SNPs. This raw vcf file was then filtered resulting in a dataset of 40 accessions (Appendix A), and 4,593 good quality SNPs out of which 3,879 (84%) SNPs were mapped by BLAST on the *D. rotundata* reference genome. The KASPar assay was then developed by selecting 192 SNPs representative of SNPs mapped along the *D. rotundata* reference sequence, and they were tested on 141 accessions.

Among the 192 SNPs, 26 (13%) SNPs failed as they did not produce any amplification signal. From the remaining 166 SNPs (87%), 129 SNPs (Appendix C) with less than 20% missing data and a minor allele frequency of over 5% were retained as high-quality SNPs. This final dataset (129 SNPs  $\times$  141 accessions) contained an overall missing data rate of only 0.5% with a maximum of 3% missing data per accession.

The 129 validated KASPar SNPs were distributed on all linkage groups used to construct the *D. rotundata* reference genome (Figure 1). Their distribution was not homogeneous along chromosomes as their position was planned to be representative of that of the initial set of 3,879 mapped SNPs and not equally spaced.

#### 3.2 | Assessment of ploidy levels

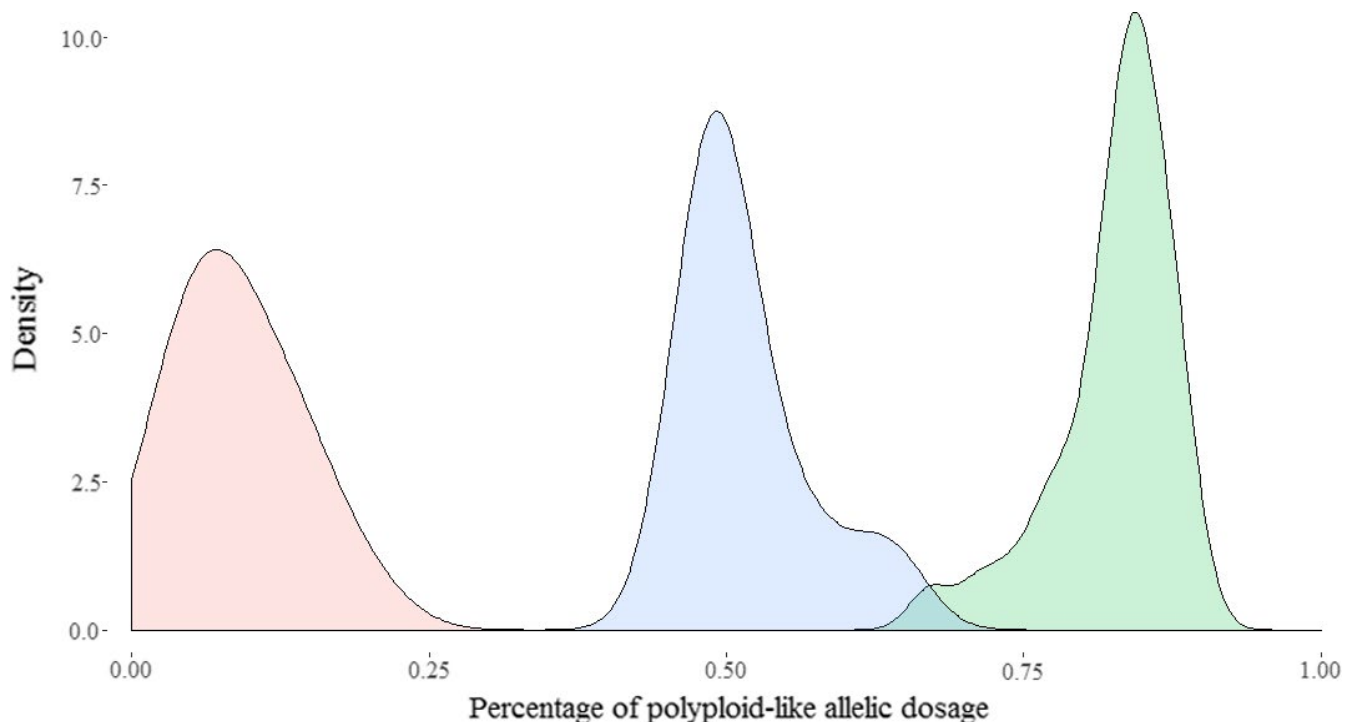
In our *D. alata* validation panel, three ploidy levels (2x, 3x and 4x) coexisted (Appendix B). Thus, the KASPar assay could theoretically produce a maximum of seven types of fluorescence signal (Table 1) corresponding to two types of fluorescence signal in homozygous states (2:0 = 3:0 = 4:0; 0:2 = 0:3 = 0:4), the fluorescence signal of mixed and balanced allelic dosages (1:1 for diploids or 2:2 for tetraploids) and the four types of fluorescence signal corresponding to the different possible unbalanced allelic dosages at heterozygotic loci ("polyploid-like" in Table 1) of triploids and tetraploids (1:3; 1:2; 2:1;



**FIGURE 1** Location of KASPar SNPs on the *D. rotundata* reference genome (Tamiru et al., 2017). The 21 linkage group are aligned from left to right. Black dots, failed or bad quality SNPs; red dots, the 129 validated SNPs

Type of genotype	Ploidy	Allelic		Type of fluorescence signal	
		Dosage	Composition	Theo.	Obs.
Diploid-like	Diploid	0:2	X:X	1	1
		1:1	X:Y	4	3
		2:0	Y:Y	7	5
	Triploid	0:3	X:X:X	1	1
		3:0	Y:Y:Y	7	5
	Tetraploid	0:4	X:X:X:X	1	1
		2:2	X:X:Y:Y	4	3
4:0		Y:Y:Y:Y	7	5	
Polyploid-like	Triploid	1:2	X:X:Y	3	2
		2:1	X:Y:Y	5	4
	Tetraploid	1:3	X:X:X:Y	2	2
		3:1	X:Y:Y:Y	6	4

**TABLE 1** Summary of genotype, allelic composition and fluorescence signals



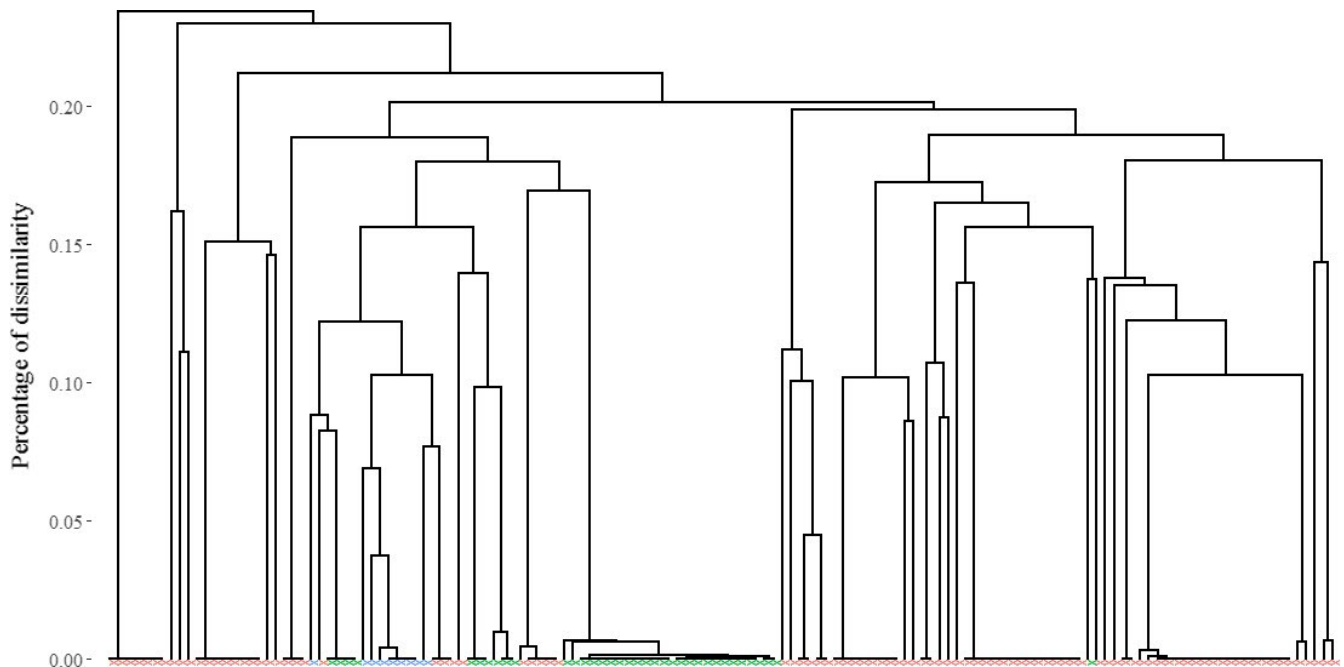
**FIGURE 2** Distribution of the percentage of polyploid-like genotypes (1:3, 1:2, 2:1, and 3:1 allelic dosage) on overall heterozygous genotypes by ploidy level (red, diploid; green, triploid; blue, tetraploid)

3:1). In our case, due to insufficient fluorescence resolution, it was not possible to distinguish fluorescence signals of the 1:3 tetraploid allelic dosage from the 1:2 triploid allelic dosage, or the 2:1 triploid allelic dosage from the 3:1 tetraploid allelic dosage. Consequently, a maximum of five types of fluorescence signals were identified. Overall, five, four, three, and two allelic dosages were detected for 64 (50%), 41 (32%), 19 (15%), and 5 (4%) SNPs, respectively, because some allelic dosages were not present in the validation panel or they were cofounded.

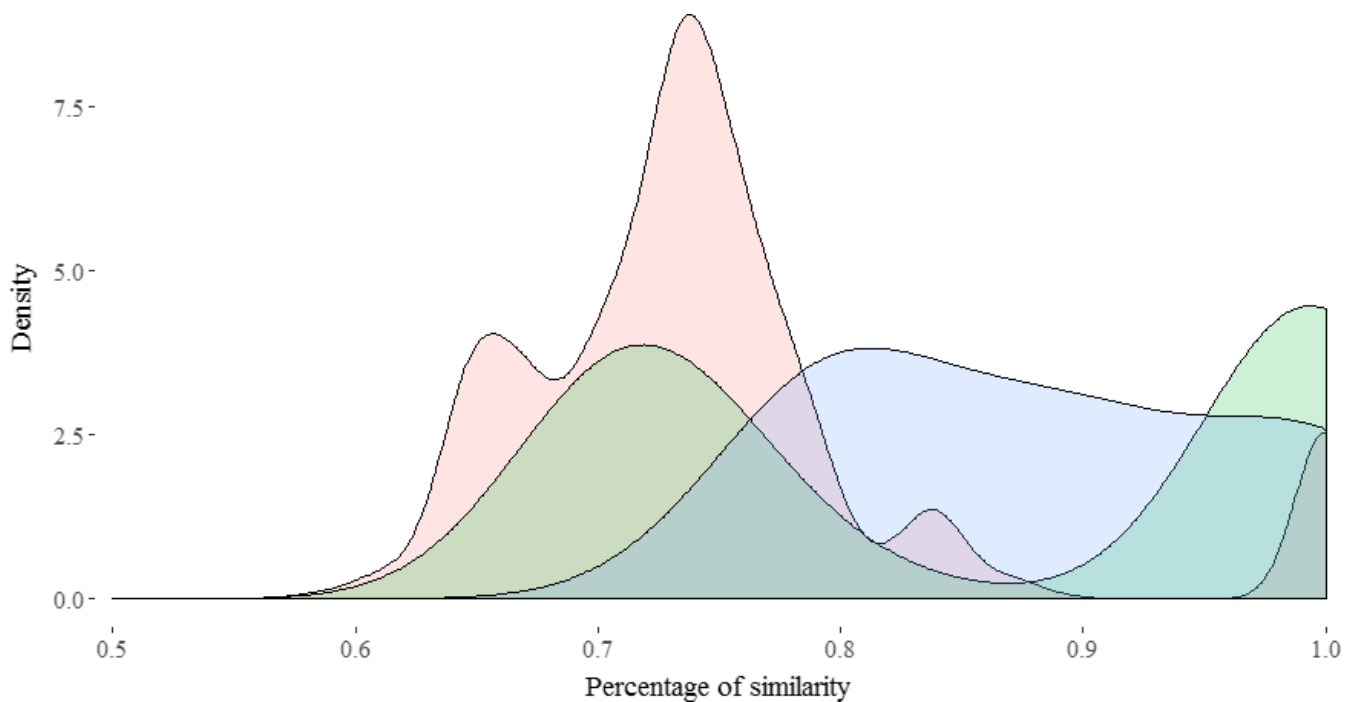
However, the overall allele call and allelic dosage assessment quality were good. Indeed, the ratio of genotypes scored as

“polyploid-like” on overall heterozygous genotypes by accession was low ( $0.09 \pm 0.05$ ) for diploids and high for triploids ( $0.83 \pm 0.05$ ). In addition, the three distributions of this ratio corresponding to the three ploidy levels did almost not overlap (Figure 2).

We were thus not able to differentiate all allelic dosage from each other when looking at one SNP. However, ploidy level could be deduced when taking all the KASPar array into account and considering the proportion of genotypes scored as “polyploid-like” per accession. This KASPar assay thus differentiated the accession ploidy level and allowed us to assign it for 12 accessions originally of unknown ploidy. Nine were set as diploid and three as triploid.



**FIGURE 3** Dendrogram of dissimilarity between 141 *D. alata* accessions (red, diploid; green, triploid; blue, tetraploid)



**FIGURE 4** Distribution of similarity between all accession pairs by ploidy (red, diploid; green, triploid; blue, tetraploid)

### 3.3 | Diversity analysis

Overall, 141 accessions from CRB-PT and CIRAD ex situ collections in Guadeloupe were used to validate the KASPar assay (96 diploids, 36 triploids, and nine tetraploids including accessions with known and deduced ploidy level).

The allele presence and/or absence was used to assess the similarity between accessions and thus to identify duplicate accessions

(Figure 3). Indeed, by defining redundancy groups, we ended up with 43 nonredundant groups each containing one to 24 accessions.

These groups of genetically similar accessions were partially expected based on the accession vernacular names. For example, the second biggest group (redundancy group 6, Appendix B) was composed of 18 accessions, five of which had a name related to “Saint Vincent.” The third biggest group contained 14 accessions, four of which had a name related to “Pacala.”

The main group of redundant accessions was composed of 24 triploids collected at several distant locations (Caribbean islands, New Caledonia and Madagascar). This group consisted of 67% (24/36) of the triploid accessions present in the CRB-PT and CIRAD collections.

More generally, redundancy groups only consisted of accessions with the same ploidy level (Figure 4). Moreover, similarities within triploids or within tetraploids were higher than within diploids.

The diversity analysis was based on these 43 redundancy groups to avoid bias. After clustering, the bootstrap procedure detected five significant gene pools, named “cluster” here, represented in the kinship network (Figure 5). Only one (cluster C, Figure 5) consisted of accessions from the three ploidy levels. This cluster encompassed accessions from the Caribbean and Pacific regions. Clusters A, B, and D contained triploids from the Caribbean and Madagascar, tetraploids from the Pacific and diploids from the Caribbean, respectively (Figure 5, Appendix B). Cluster E was the biggest one, with 21 nonredundant diploid accessions originating from India, Nigeria, Côte d'Ivoire, the Caribbean and Pacific (Figure 5, Appendix B).

Genotype permutations and network analysis gave a more detailed view of kinship between redundancy groups and Clusters. This approach revealed a low number of significant links between the diversity clusters D or E and the others (Figure 5) revealing that these clusters could consist of original gene pools.

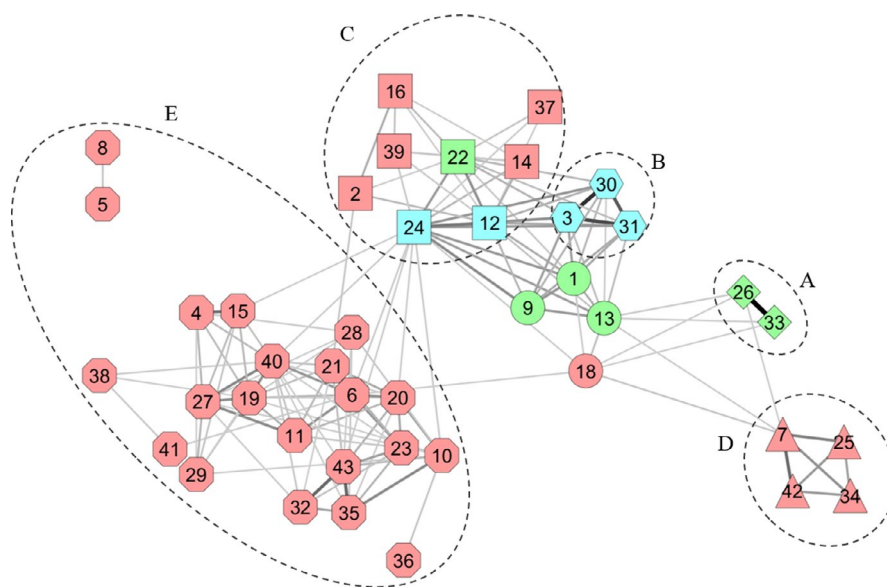
## 4 | DISCUSSION

### 4.1 | Assessment of allelic dosage and detection of ploidy levels

KASPar technology is based on competitive allele-specific amplification followed by allele-specific fluorescence assessment

(Semagn, Babu, Hearne, & Olsen, 2014). Detection of allelic dosage in polyploid species is thus possible (Cuenca, Aleza, Navarro, & Ollitrault, 2013). However, several parameters may influence the fluorescence, such as the DNA quality or primer specificity, and consequently the ability to discriminate fluorescence signals and the allelic dosage. In our case, we were able to discriminate five types of fluorescence signal. At heterozygous loci, fluorescence signals were a mixture of two types of allelic-specific fluorescence. Fluorescence signals should also be balanced for diploids which have a balanced allelic dosage (1:1) at heterozygous loci. Diploids should therefore theoretically have no genotypes assessed as “polyploid-like.” Conversely, triploids should theoretically have only genotypes assessed as “polyploid-like” at heterozygous loci. A balanced allelic dosage is impossible for triploids. Our results showed that  $91 \pm 5\%$  and  $83 \pm 5\%$  of heterozygous genotypes were correctly called for diploids and triploids, respectively. Regarding the recent explosion of genotyping related to next-generation sequencing, bioinformatics tools have been developed to accurately determine dosages (e.g., GBS2ploidy; Gompert & Mock, 2017). However, this requires deep sequencing and usually an assumption of ploidy levels present in the dataset (Bourke, Voorrips, Visser, & Maliepaard, 2018).

Application in collection management may nevertheless not require allelic dosage assessment at each locus. Our aim was thus to develop a tool for estimating ploidy levels and not variations in copy number. Moreover, the results showed that ploidy levels for each accession can be accurately deduced from the percentage of “polyploid-like” genotypes on overall heterozygous genotypes. Regarding the overlapping distributions of this ratio (Figure 2), the only risk is to confuse triploids and tetraploids estimated at 3%. Consequently, ploidy level assessment is possible and fairly accurate for *D. alata* using the KASPar assay developed in this study.



**FIGURE 5** Network of kinship for the 43 *D. alata* redundancy groups based on significant similarity ( $p < 0.05$ , edge-weighted spring-embedded layout). Nodes shape and letter, cluster of diversity identified by a bootstrap procedure; red nodes, diploids; green nodes, triploids; blue nodes, tetraploids; edge colors, similarity from gray (0.64) to black (1)

## 4.2 | Identification of duplicate accessions

The dataset included 129 SNPs validated on 141 accessions corresponding to 43 unique redundancy groups. The resuming of the 141 accessions to 43 unique redundancy groups was related to the narrow *D. alata* genetic diversity, above all in polyploid germplasm (i.e., triploids and tetraploids) already identified in previous studies. For example, using DaRT markers, a low varietal richness was revealed by Vandenbroucke et al. (2016), who studied 80 landraces from six different Vanuatu islands and differentiated only seven unique genotypes. Using isozyme markers, Lebot et al. (1998) studied 269 worldwide distributed cultivars and concluded that the genetic diversity of the most widespread cultivars was narrow.

Regarding the accession vernacular names, redundant accessions were expected in our sample. Some of these redundancy groups contained accessions detected in duplicate, while they could be differentiated by morphological characterization. For example, redundancy group five (including Lupias, Malalagi, or Malankon) exhibited diversity in tuber shape and tuber flesh color in agreement with previous genetic diversity studies that already pooled these accessions together and highlighted this intragroup variability in tubers (Arnau et al., 2017; Malapa et al., 2005).

Morphological variability within a redundancy cluster mostly arises via *D. alata* clonal reproduction and farmers' selection of new morphotypes resulting from somatic mutations (Lebot et al., 1998; Malapa et al., 2005; Vandenbroucke et al., 2016). Small genetic or epigenetic variations are commonly selected to create new diversity in horticultural crops such as yam as reviewed by Krishna et al. (2016).

The ability of KASPar assay developed in this study to differentiate duplicates in collections from genetically close accessions was related: (a) to the low number of studied loci (129), but also (b) to the *D. alata* diversification process (i.e., selection of somaclonal mutants) and (c) the presence of real duplicates within collections. This tool is thus efficient for attributing accessions to a genetic lineage (e.g., germplasm exchange), but a good complementary agro-morphological and ecophysiological characterization of collections should also be done to completely differentiate somaclonal mutant clones from duplicates (e.g., identification of promising genitors for breeding programs).

## 4.3 | Diversity and collection management

The CRB-PT collection has been shown to be representative of worldwide *D. alata* diversity (Arnau et al., 2017). A subset of this ex situ collection has been genotyped in this study. However, all diversity groups identified by Arnau et al. (2017) were present (except one containing five very similar Indian accessions). Our validation panel was thus representative of the worldwide *D. alata* diversity. Moreover, a good correlation was obtained between the findings of the previous study of worldwide *D. alata* diversity of Arnau et al. (2017) and the gene pools identified in this study (Appendix B). We can thus hypothesize that the 129 SNPs KASPar array developed for *D. alata* allow us to accurately assess genetic diversity and the

findings may be transferable to other collections. Moreover, this genotyping tool is a robust method: (a) to assess complementarity/redundancy between the different collections, (b) to identify under represented genetic groups, and (c) to plan future collects to fill gaps in collections.

## 5 | CONCLUSION

This is the first SNP array designed for *D. alata* and validated on a subset of accessions representative of worldwide *D. alata* diversity. This tool will allow users to estimate accession ploidy levels and genetic lineages. The results showed a good correlation between the diversity assessed by this KASPar array and the findings of previous studies. This KASPar array is a robust and cost-effective tool for diversity assessment and collections management. Regarding the importance of vegetative reproduction and somaclonal selection in *D. alata*, it is a good tool to complement agro-morphological description in collections.

## ACKNOWLEDGMENTS

This study was financially supported by the European Union and Guadeloupe Region (Programme Opérationnel FEDER—Guadeloupe—Conseil Régional 2014–2017). The authors would like to thank Suzia Gélabale, Marie-Claire Gravillon, Jean-Luc Irep, David Lange, and Elie Nudol for their involvement in CRB-PT and CIRAD in vitro and field collections conservation. Finally, we are grateful to Patrick Ollitrault for his valuable discussion and to David Manley for English proofing.

## CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

## AUTHOR CONTRIBUTIONS

C.P., F.C., H.C., and P.M. designed the study. C.P., F.C., E.M., G.A., and R-M.G. contributed to collecting materials and sample preparation. P.M. and S.C. developed GBS protocol, carried out DNA extraction, and GBS library preparation. H.C. and P.M. performed SNP discovery. F.C. and H.C. designed the KASPar assay and performed its analysis. C.P., F.C., and H.C. wrote the manuscript with the input of all authors.

## DATA ACCESSIBILITY

Plant materials may be requested at the CRB-PT of Guadeloupe <http://intertrop.antilles.inra.fr/Portail/accessions/find/11>. KASPar primers sequence is available in Appendix B.

## ORCID

Fabien Cormier  <https://orcid.org/0000-0002-7283-4616>

## REFERENCES

- Arnau, G., Bhattacharjee, R., Mn, S., Chair, H., Malapa, R., Lebot, V., ... Pavis, C. (2017). Understanding the genetic diversity and population structure of yam (*Dioscorea alata* L.) using microsatellite markers. *PLoS One*, 12(3), e0174150.
- Arnau, G., Némorin, A., Maledon, E., & Abraham, K. (2009). Revision of ploidy status of *Dioscorea alata* L. (Dioscoreaceae) by cytogenetic and microsatellite segregation analysis. *Theoretical and Applied Genetics*, 118, 1239–1249. <https://doi.org/10.1007/s00122-009-0977-6>
- Asemota, H. N., Ramser, J., Lopez-Peralta, C., Weising, K., & Kahl, G. (1996). Genetic variation and cultivar identification of Jamaican yam germplasm by random amplified polymorphic DNA analysis. *Euphytica*, 92, 341–351. <https://doi.org/10.1007/BF00037118>
- Bourke, P. M., Voorrips, R. E., Visser, R. G. F., & Maliepaard, C. (2018). Tools for genetic studies in experimental populations of polyploids. *Frontiers in Plant Science*, 9, 513. <https://doi.org/10.3389/fpls.2018.00513>
- Broccanello, C., Chiodi, C., Funk, A., McGrath, J. M., Panella, L., & Stevanato, P. (2018). Comparison of three PCR-based assays for SNP genotyping in plants. *Plant Methods*, 14, 28. <https://doi.org/10.1186/s13007-018-0295-6>
- Chair, H., Sardos, J., Supply, A., Mournet, P., Malapa, R., & Lebot, V. (2016). Plastid phylogenetics of Oceania yams (*Dioscorea* spp., Dioscoreaceae) reveals natural interspecific hybridization of the greater yam (*D. alata*). *Botanical Journal of the Linnean Society*, 180, 319–333.
- Cormier, F., Lawac, F., Maledon, E., Gravillon, M.-C., Nudol, E., Mournet, P., ... Arnau, G. (2019). A reference high-density genetic map of greater yam (*Dioscorea alata* L.). *Theoretical and Applied Genetics*. <https://doi.org/10.1007/s00122-019-03311-6>
- Cuenca, J., Aleza, P., Navarro, L., & Ollitrault, P. (2013). Assignment of SNP allelic configuration in polyploids using competitive allele-specific PCR: Application to citrus triploid progeny. *Annals of Botany*, 111, 731–742. <https://doi.org/10.1093/aob/mct032>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Davey, J. W., Hohenlohe, P., Etter, P., Boone, J., Catchen, J., & Blaxter, M. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12, 499–510. <https://doi.org/10.1038/nrg3012>
- Dereeper, A., Nicolas, S., Lecunff, L., Bacilieri, R., Doligez, A., Peros, J. P., ... This, P. (2011). SNIPlay: a web-based tool for detection, management and analysis of SNPs. Application to grapevine diversity projects. *BMC Bioinformatics*, 12, 134.
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 6, e19379. <https://doi.org/10.1371/journal.pone.0019379>
- García-Lor, A., Ancillo, G., Navarro, L., & Ollitrault, P. (2013). Citrus (Rutaceae) SNP markers based on Competitive Allele-Specific PCR; transferability across the Aurantioideae subfamily. *Applications in Plant Sciences*, 1(4), apps.1200406. <https://doi.org/10.3732/apps.1200406>
- Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q. I., & Buckler, E. S. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS One*, 9(2), e90346. <https://doi.org/10.1371/journal.pone.0090346>
- Gompert, Z., & Mock, K. E. (2017). Detection of individual ploidy levels with genotyping-by-sequencing (GBS) analysis. *Molecular Ecology Resources*, 17, 1156–1167. <https://doi.org/10.1111/1755-0998.12657>
- Hiremath, P. J., Kumar, A., Penmetsa, R. V., Farmer, A., Schlueter, J. A., Chamarthi, S. K., ... Varshney, R. K. (2012). Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant Biotechnology Journal*, 10, 716–732. <https://doi.org/10.1111/j.1467-7652.2012.00710.x>
- IPGRI/IITA (1997). Descriptors for Yam (*Dioscorea* spp.). International Institute of Tropical Agriculture, Ibadan, Nigeria/International Plant Genetic Resources Institute, Rome, Italy.
- Krishna, H., Alizadeh, M., Singh, D., Singh, U., Chauhan, N., Eftekhari, M., & Sadh, R. K. (2016). Somaclonal variations and their applications in horticultural crops improvement. *3 Biotech*, 6, 54. <https://doi.org/10.1007/s13205-016-0389-7>
- Langmead, B., & Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357–359. <https://doi.org/10.1038/nmeth.1923>
- Lebot, V., Trilles, B., Noyer, L. J., & Modesto, J. (1998). Genetic relationships between *Dioscorea alata* L. cultivars. *Genetic Resources and Crop Evolution*, 45, 499–509.
- Mahalakshmi, V., Ng, Q., Atalobor, J., Oguniola, 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>
- Malapa, R., Arnau, G., Noyer, J. L., & Lebot, V. (2005). Genetic diversity of the greater yam (*Dioscorea alata* L.) and relatedness to *D. nummularia* Lam. and *D. transversa* Br. as revealed with AFLP markers. *Genetic Resources and Crop Evolution*, 52, 919–929. <https://doi.org/10.1007/s10722-003-6122-5>
- Nemorin, A., David, J., Maledon, E., Nudol, E., Dalon, J., & Arnau, G. (2013). Microsatellite and flow cytometry analysis to help understand the origin of *Dioscorea alata* polyploids. *Annals of Botany*, 112, 811–819. <https://doi.org/10.1093/aob/mct145>
- Otoo, E., Anokye, M. L., Asare, P. A., & Telleh, J. P. (2015). Molecular categorization of some water yam (*Dioscorea alata* L.) germplasm in Ghana using microsatellites (SSR) markers. *Journal of Agricultural Science* 7(10), 226–238.
- R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Risterucci, A.-M., Hippolyte, I., Perrier, X., Xia, L., Caig, V., Evers, M., ... Glaszmann, J.-C. (2009). Development and assessment of diversity arrays technology for highthroughput DNA analyses in Musa. *Theoretical and Applied Genetics*, 119, 1093–1103. <https://doi.org/10.1007/s00122-009-1111-5>
- 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>
- Saski, C. A., Bhattacharjee, R., Scheffler, B. E., & Asiedu, R. (2015). Genomic resources for water yam (*Dioscorea alata* L.): Analyses of EST-sequences, de novo sequencing and GBS libraries. *PLoS One*, 10(7), e0134031.
- Saxena, R. K., von Wettberg, E., Upadhyaya, H. D., Sanchez, V., Songok, S., Saxena, K., ... Varshney, R. K. (2014). Genetic diversity and demographic history of *Cajanus* spp. illustrated from genome-wide SNPs. *PLoS One*, 9, e88568.
- Semagn, K., Babu, H., Hearne, S., & Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. *Molecular Breeding*, 33, 1–14. <https://doi.org/10.1007/s11032-013-9917-x>
- Semagn, K., Beyene, Y., Makumbi, D., Mugo, S., Prasanna, B. m., Magorokosho, C., & Atlin, G. (2012). Quality control genotyping for assessment of genetic identity and purity in diverse tropical maize inbred lines. *Theoretical and Applied Genetics*, 125, 1487–1501. <https://doi.org/10.1007/s00122-012-1928-1>

- Siqueira, M. V., Marconi, T. G., Bonatelli, M. L., Zucchi, M. I., & Veasey, E. A. (2011). New microsatellite loci for water yam (*Dioscorea alata*, Dioscoreaceae) and cross-amplification for other *Dioscorea* species. *American Journal of Botany*, 98, 144–146.
- Su, T., Li, P., Yang, J., Sui, G., Yu, Y., Zhang, D., ... Zhang, F. (2018). Development of cost-effective single nucleotide polymorphism marker assays for genetic diversity analysis in *Brassica rapa*. *Molecular Breeding*, 38, 42. <https://doi.org/10.1007/s11032-018-0795-0>
- Suzuki, R., & Shimodaira, H. (2006). Pvcust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 12, 1540–1542. <https://doi.org/10.1093/bioinformatics/btl117>
- Tamiru, M., Natsume, S., Takagi, H., White, B., Yaegashi, 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, 86. <https://doi.org/10.1186/s12915-017-0419-x>
- Vandenbroucke, H., Mournet, P., Vignes, H., Chair, H., Malapa, R., Duval, M. F., & Lebot, V. (2016). Somaclonal variants of taro (*Colocasia esculenta* Schott) and yam (*Dioscorea alata* L.) are incorporated into farmers' varietal portfolios in Vanuatu. *Genetic Resources and Crop Evolution*, 63, 495–511. <https://doi.org/10.1007/s10722-015-0267-x>

**How to cite this article:** Cormier F, Mournet P, Causse S, et al. Development of a cost-effective single nucleotide polymorphism genotyping array for management of greater yam germplasm collections. *Ecol Evol.* 2019;9:5617–5636. <https://doi.org/10.1002/ece3.5141>

## APPENDIX A

**TABLE A1** Description of the 40 *D. alata* accessions used to detect polymorphic SNP

Collection	Code	Name	Origin	Ploidy
CRB-PT	PT-IG-00002	Pakutrany	Nlle Caledonie	
	PT-IG-00006	Fénakué	Puerto Rico	2
	PT-IG-00010	Divin 1	Guadeloupe	2
	PT-IG-00020	DA 26	Guyane Fr	3
	PT-IG-00338	HYB 30	Guadeloupe	
	PT-IG-00350	Pacala	Guadeloupe	2
	PT-IG-00029	Plimbite	Haïti	2
	PT-IG-00033	Pyramide	Puerto Rico	2
	PT-IG-00046	Sea 190	Puerto Rico	2
	PT-IG-00053	Kokoéta	Nlle Calédonie	2
	PT-IG-00686	Roujol		4
	PT-IG-00687	INRA C 143		
	PT-IG-00688	INRA AL 56		
	PT-IG-00690	INRA AL 18		
	PT-IG-00692	INRA X 154	Guadeloupe	
	PT-IG-00693	INRA X 17	Guadeloupe	
	PT-IG-00694	Dou		4
	PT-IG-00695	INRA X 142	Guadeloupe	
	PT-IG-00696	Ciradienne		4
	PT-IG-00697	TiViolet		4
	PT-IG-00698	Malalagi	Vanuatu	2
	PT-IG-00702	Manlankon	Vanuatu	2
	PT-IG-00689	Nureangdan	Vanuatu	3
	PT-IG-00077	Kinabayo	Puerto Rico	2
	PT-IG-00078	Toro	Haïti	3

(Continues)

**TABLE A1** (Continued)

Collection	Code	Name	Origin	Ploidy
Cirad	Vu 024a	Tépuva	Vanuatu	2
	Vu 528a	Tacharamivar		2
	Vu 564a	Mendrovar	Vanuatu	2
	Vu 567a	Homb	Vanuatu	2
	Vu 754a	Intejegan	Vanuatu	4
	Vu 231a	Tagabé	Vanuatu	4
	Ovy taty		Madagascar	
	Vu 247a	n.a	Vanuatu	2
	Vu 401a	Basa	Vanuatu	2
	Kabusa			2
	74F			2
	42F			2
	61F			2
	14M			2
	H4x200			4

**APPENDIX B****TABLE B1** Description of the 141 *D. alata* used as the KASPar assay validation panel

Collection Code	Ploidy <sup>a</sup>	Div. Clust. <sup>b</sup>	Redund. Grp <sup>c</sup>	Accession name	Origin	SSR <sup>d</sup>
PT-IG-00087	3	A	26	65	Martinique	XII
PT-IG-00070	3	A	26	66	Martinique	XII
PT-IG-00090	3	A	26	Caillade 1	Haïti	XII
PT-IG-00020	3	A	26	DA 26	French Guyana	XII
PT-IG-00037	3	A	26	DA 27	French Guyana	XII
PT-IG-00022	3	A	26	De agua	Puerto Rico	XII
PT-IG-00061	3	A	26	Igname d eau	Martinique	XII
PT-IG-00550	3	A	26	Montpellier		XII
PT-IG-00075	3	A	26	Renta Yam	Jamaica	XII
PT-IG-00072	3	A	26	Sassa 1	Martinique	XII
PT-IG-00063	3	A	26	Sassa 2	Martinique	
PT-IG-00088	3	A	26	St Martin	Martinique	XII
PT-IG-00034	3	A	26	Sweet yam	Jamaica	XII
PT-IG-00557	3	A	26	Tahiti couleuvre	Guadeloupe	XII
PT-IG-00068	3	A	26	Tahiti cultivé	Guadeloupe	XII
PT-IG-00069	3	A	26	Tahiti French	Guadeloupe	XII
PT-IG-00018	3	A	26	Tahiti messien	Guadeloupe	
PT-IG-00064	3	A	26	Tana	New Caledonia	XII
PT-IG-00021	3	A	26	Telemaque	Martinique	XII
PT-IG-00044	3	A	26	Ti Joseph 1	Haïti	XII
PT-IG-00078	3	A	26	Toro	Haïti	XII
CT257_CIV	3	A	26	OvyTaty AmbalaKindresy-Ambohimaso	Madagascar	
CT258_CIV	3	A	26	OvyTaty Amboasary-Ambohimaso	Madagascar	

(Continues)

TABLE B1 (Continued)

Collection Code	Ploidy <sup>a</sup>	Div. Clust. <sup>b</sup>	Redund. Grp <sup>c</sup>	Accession name	Origin	SSR <sup>d</sup>
PT-IG-00685	3	A	26	Sainte Anne		
PT-IG-00030	3	A	33	67	Martinique	XII
PT-IG-00558	4	B	3	Wabé	New Caledonia	XVIII
Vu472a	4	B	3	Toufi Tetea	Vanuatu	XVIII
Vu231a	4	B	3		Vanuatu	XVIII
Vu750a	4	B	3	Wanorak	Vanuatu	
Vu534a	4	B	3	Bisoro	Vanuatu	XVIII
Vu754a	4	B	30	Noulelcae	Vanuatu	XVI
Vu408a	4	B	31	Manioc	Vanuatu	
PT-IG-00039	2	C	2	Americano	Dominican Republic	VII
PT-IG-00023	2	C	2	Florida	Puerto Rico	
PT-IG-00553	2	C	2	Pro 1		VII
PT-IG-00095	2	C	2	SEA 144	Puerto Rico	IV
PT-IG-00555	2	C	2	SRT 29		VII
PT-IG-00041	2	C	2	St Domingue	Dominican Republic	VII
Vu401a	2	C	2	Basa	Vanuatu	VII
CT256	2	C	2			
PT-IG-00009	4	C	12	Nouméa	New Caledonia	XVI
Vu247a	2	C	14		Vanuatu	
Vu528a	2	C	16	Sinoua	Vanuatu	
PT-IG-00025	3	C	22	Goana	New Caledonia	XIII
PT-IG-00002	3	C	22	Pakutrany	New Caledonia	XIII
Vu699a	3	C	22	Tumas	Vanuatu	
Vu461a	3	C	22	Tumas	Vanuatu	XIII
Vu755a	4	C	24	Nepelev	Vanuatu	
PT-IG-00014	2	C	37	Divin 2	Guadeloupe	
PT-IG-00006	2	C	37	Fénakué	Puerto Rico	
PT-IG-00053	2	C	37	Kokoéta	New Caledonia	
PT-IG-00559	2	C	39	Wassa	New Caledonia	
PT-IG-00001	2	D	7	64	Martinique	
PT-IG-00010	2	D	7	Divin 1	Guadeloupe	
PT-IG-00568	2	D	25	77	Martinique	IV
PT-IG-00092	2	D	34	Caplaou	Puerto Rico	
PT-IG-00561	2	D	42	H 23		
PT-IG-00562	2	D	42	H 50		
74F	2	E	4		India	
PT-IG-00049	2	E	5	Cinq	Puerto Rico	III
PT-IG-00027	2	E	5	Lupias	New Caledonia	III
PT-IG-00046	2	E	5	Sea 190	Puerto Rico	III
Vu590a	2	E	5		Vanuatu	III
Vu423a	2	E	5	Manlankon	Vanuatu	III
Vu639a	2	E	5	Malalagi	Vanuatu	III
Vu024a	2	E	5	Ptris	Vanuatu	III
PT-IG-00065	2	E	6	DA 28	French Guyana	IV

(Continues)

TABLE B1 (Continued)

Collection Code	Ploidy <sup>a</sup>	Div. Clust. <sup>b</sup>	Redund. Grp <sup>c</sup>	Accession name	Origin	SSR <sup>d</sup>
PT-IG-00093	2	E	6	DA 32		
PT-IG-00395	2	E	6	Fafadro bis		IV
PT-IG-00060	2	E	6	Grand Etang	Guadeloupe	IV
PT-IG-00051	2	E	6	Morado	Cuba	IV
PT-IG-00073	2	E	6	Purple Lisbon	Puerto Rico	IV
PT-IG-00333	2	E	6	Sainte Catherine	Guadeloupe	IV
PT-IG-00052	2	E	6	Smooth Statia	Puerto Rico	IV
PT-IG-00024	2	E	6	St Vincent blanc 1	Martinique	IV
PT-IG-00036	2	E	6	St Vincent blanc 2	Martinique	IV
PT-IG-00556	2	E	6	St Vincent mart.	Guadeloupe	IV
PT-IG-00045	2	E	6	St Vincent Violet	Martinique	IV
PT-IG-00016	2	E	6	St Vincent Yam	St. Lucia	IV
PT-IG-00374	2	E	6	Ti Joseph	Haïti	IV
PT-IG-00067	2	E	6	Wénéféla bis	New Caledonia	IV
Vu487a	2	E	6	Teroosi	Vanuatu	VI
770	2	E	6			
PT-IG-00623	2	E	6			
PT-IG-00396	2	E	8	A 24		
PT-IG-00071	2	E	10	72	Martinique	VIII
PT-IG-00055	2	E	10	76	Martinique	VIII
PT-IG-00089	2	E	10	Asmhore		
PT-IG-00058	2	E	10	Bété Bété	Côte d'Ivoire	VIII
PT-IG-00091	2	E	10	Campêche 2		
PT-IG-00546	2	E	10	Jardin Haitien		VIII
PT-IG-00547	2	E	10	Kourou 1	French Guyana	VIII
PT-IG-00548	2	E	10	Kourou 2	French Guyana	VIII
PT-IG-00350	2	E	10	Pacala	Guadeloupe	VIII
PT-IG-00551	2	E	10	Pacala cacao	French Guyana	VIII
PT-IG-00552	2	E	10	Pacala Guyane	French Guyana	VIII
PT-IG-00017	2	E	10	Pacala station	Guadeloupe	VIII
PT-IG-00554	2	E	10	SRT 24		VIII
19	2	E	10			
PT-IG-00057	2	E	11	Vino Purple forme	Puerto Rico	
61F	2	E	15		India	
PT-IG-00019	2	E	19	Gordito	New Caledonia	IX
PT-IG-00047	2	E	20	Buet	New Caledonia	
PT-IG-00029	2	E	20	Plimbite	Haïti	
PT-IG-00048	2	E	21	Bacala 1	Haïti	
PT-IG-00413	2	E	21	St Vincent	St. Vincent	
Cuba6	2	E	23		Cuba	
PT-IG-00542	2	E	27	AL 10		I
PT-IG-00042	2	E	27	Brazzo Fuerte	Puerto Rico	I
PT-IG-00038	2	E	27	Brésil 1		I
PT-IG-00564	2	E	27	KL 10		I
PT-IG-00565	2	E	27	KL 21		

(Continues)

TABLE B1 (Continued)

Collection Code	Ploidy <sup>a</sup>	Div. Clust. <sup>b</sup>	Redund. Grp <sup>c</sup>	Accession name	Origin	SSR <sup>d</sup>
PT-IG-00566	2	E	27	KL 40		I
PT-IG-00054	2	E	27	MP1 16H56		I
PT-IG-00033	2	E	27	Pyramide	Puerto Rico	I
PT-IG-00074	2	E	28	Oriental	Barbados	II
14M	2	E	29		India	
PT-IG-00077	2	E	32	Kinabayo	Puerto Rico	II
PT-IG-00085	2	E	35	St Sauveur	Guadeloupe	
PT-IG-00560	2	E	35	Yam jamaïque		
PT-IG-00543	2	E	36	Cross lisbon		
PT-IG-00392	2	E	38	A 13		
PT-IG-00398	2	E	38	A 2		
PT-IG-00563	2	E	40	Sc.c 1.1		
PT-IG-00008	2	E	41	AIA 445	Nigeria	
PT-IG-00015	2	E	43	Igname rouge	Guadeloupe	X
Vu703a	3	F	1	Nawanurunkimanga	Vanuatu	
PT-IG-00544	3	F	9	Cuello largo	Puerto Rico	XV
PT-IG-00026	3	F	9	Féo	Puerto Rico	XV
Vu696a	3	F	9	Nowateknempian	Vanuatu	XV
PT-IG-00076	3	F	13	Bélep	New Caledonia	XIV
Vu735a	3	F	13	Noplou	Vanuatu	XIV
Vu760a	3	F	13	Nureangdan	Vanuatu	XIV
PT-IG-00397	2	F	17	SEA 119, Toki		
Vu613a	2	F	17	Peter	Vanuatu	VI
Vu589a	2	F	17	Makila	Vanuatu	XI
VU590a	2	F	18		Vanuatu	III
Vu554a	2	F	18	Nourembor	Vanuatu	VI
Vu567a	2	F	18	Letsletsbolos	Vanuatu	IV
Vu564a	2	F	18	Makila	Vanuatu	VI
Vu026a	2	F	18	Dammasis	Vanuatu	VI

<sup>a</sup>In italic, ploidy detected using the percentage of polyploid genotype type on overall heterozygous loci. <sup>b</sup>Group of diversity from diversity analysis. <sup>c</sup>Group of similarity used to select nonredundant accessions. Genotypes in the same group have a maximum of one allele mismatch). <sup>d</sup>Cluster of diversity identified by SSR in Arnau et al. (2017).



TABLE C1 (Continued)

SNP_ID	# Fluo.type	Chr.	Pos.	E-value	Sequence
S1_94822591	5	4	6518043	4E-37	TCGTGATCGTGTCTCTCATCAAAATGTTAGATGTTGTTCTAATCTTCAAGCAATCA[G/A] GAACCTATTTGCTACTATGAGTTGTGACTTATTTGTTGCTGGTCACTGGATACTGCAGGTC
S1_223059854	5	4	9590872	4E-49	GTTGTGCCAGCAATCATGAGACCATTATGAGCTATGTTGATGGATGGTGGGTCGGG[A/C/T] GTCTGTGGATAAGCTGTAACGTCAGGCAGCTACCACCATTGGAGAAAATGGCCAGCT
S3_1940455	4	4	11602024	4E-49	ATGCCTGAATCTGGAGGACAAGCTACTGCAGTGTCAAGAAAATAGACATACTTGAAGAG[A/C] ATTATGAATCAGAACAGTCCAGGCGGTAATGTTGAGTTTCATGCTTCTGTTCCTTCT
S2_59089982	4	4	13022373	2E-47	CTCAATAATGGTGGACAAAATGTCCTTCTAAATTCAGAAAAAAAATGTCTACTAATACCT[C/G] TCCAGCTCTGCATTTACTTGTATAGGATGACATAATGATGACTCTGCAGAGGATGAT
S1_29508975	5	4	13956140	8E-52	TAGTGACAGTTGAGATCAITTAACGAGCATAGAAGAACTACTGCAGAAAGCCAAAAGCC[A/T] GAAGTCAATGGAGTTCTATGGAACATAATGAGAGGAATTTGGAAACACTATATTTGCACCT
S1_78099239	4	4	19009721	2E-33	CAGGTTTTCTTTTCAATTCGAAGAGACATCAAGCAAAAGGCTTGCAGAAAACCGATACCAAA[C/G] CTGAGGTAGTATGCTTATCATTGTTGATAAAATTCAGTAAACCTGCAGGCAATTAGTAATG
S1_98182899	5	4	25059970	2E-47	TGTGACAATGCTATGACCACACATATGTTTCAAACAGACAAGGAACAAGAAATCAACAGA[T/C] ACACTACTTACAGTGACAAGCTCCGATATCAGTTGCGCAGACAAACCTGTTTTCTGCAGC
S1_13376874	3	4	25692581	2E-40	CTTGATATGCTGTATTGGGTGTCATTTCTGCAGTATTATTGGTCCGCAAGGTAATCA[A/G] ATAGAAAATTTGGCGTCAITTAGCAATATCAAACCTGATTTTGAATTTGTAGTTTCGAGCATT
S1_50863270	5	4	27382503	2E-54	GTTGGCTCAAAGACTGAGGAAGTCCGTGAAGATAGGGCTCATTGGGTACCATAATAAC[T/C] GGTGATATTATGGTCAGGGTTGGATCAGTGAATATGATGGATATTCATTTGGTGTCTGCA
S1_30240813	5	5	4418552	4E-43	TCAGAGAGTGATCATAAAAAAATAGATTTCATTTGTCATTGTGATGTCCTCCCTGATCTTT[A/C] TTACTTGCCCCCAAACGATAAGGACATCTTTCTCCACATAACTTGTCAATAATTTCTGCAGA
S1_284257251	5	5	8809466	2E-41	AATTTTCTGATCTGAGTATGGTCAACAAGAATCCAAACATAAACTCAAGTAAATGCAGT[C/C] AAATFACAATTTTACATAATTTCTCCACATAACTTGTCAATAATTTCTGCAGA
S1_172013713	2	5	15915339	5E-48	AACGCATGATACTCAATGTTGTTACTAATTAATGAACTCAATTAATGACCTGCAGTTG[C/T] TTGAATTTTCATGCTATGTTTGAAGGCTCTAGTGTGCCAAAACCTCAGACATCTCCG
S2_46046547	5	5	18636385	5E-54	CACGAGGGGCTGCTTCCCTTGATGATGAAACCCAAAGAACTTATTTCTCAAATAAAGCG[C/A] TTCATTGGAAAAGAAATCTCTGACCCGGAGCTCAGCTGACCTGCAGTTATTCCTTTT
S1_161404508	4	5	19741536	5E-42	TGCTAATCAAAACTGATGCTCTGCAGGACCAAGTTGATCAATGAAAAGATCAGGGATT[A/C] TCACCTTCTATCACCTGTAGAGTACTGCATGTACATAAGTCACCATGAAAAGCACCGGGC
S1_86749745	5	5	20784276	1E-48	CCCTTCTGATATTTGCTGGAGTTGAGATGCTGTTGAACTTTAGCTGGAAAATTTACAG[T/C] CAACTCTATGAATTTGTTTCTGATTCATACCGCACATTTGTGATTTTGTGCTGCAGAG
S1_349430697	5	5	22888769	8E-52	TCCATCTGGGAAGCACGTCGCAATGTTGATTTCCGTTAGGCCAAGATTTGGAGCCCAAG[C/A] ACATCTTAAACCCCAATCGGAATGCATCTGCAGGCGAGAAAAGCATCCCAATTTCCAGCTC
S3_43353096	5	5	24048351	8E-52	TGATGGGGAAATAACCCAAACTGGTGGAGTTCTATCAGCAACATGAAGCCAACTGCAG[G/C] AGATGAAAACCTCTCTTCTTATCCTTTTCCATCTTCCATCTTCCATCTCTCCATCACTACTC
S1_51060666	3	5	26180154	1E-42	GAAGGAAGGCAGCAGCCTTCAAACCTGCAGATGAAGTCCCAAGTCTTTTGAATAATGCA[A/T] AACCCAGCCAAATTTGCAGCTGCTAGTTATAGGTAATGATAGATAGTATCTAGGCTAT
S1_126396048	5	5	27405531	4E-49	CAACTGTCTGAGAAAAGAAACTGACCCGTCATCAACAAAACAGAAATTTAGAG[G/A] AATGCCAAGGCACCTTCTCAAGGTTTCTAATCTGTTCTTGAGGGTGGCTTGCTCT

(Continues)

TABLE C1 (Continued)

SNP_ID	# Fluo.type	Chr.	Pos.	E-value	Sequence
S1_189523417	3	5	30765836	4E-49	TTTAAAGCTTGACAGCGGAGTTTGAAGACCCTGTGTCAGGCGCATCCAGCTTGGCG[A/G] GCAGGAACAGCTACACCGGCTGTACTGCCGTGGCGTTTCACTCAGTGGACTGAAAAACA
S1_210690510	5	5	31577924	2E-53	ATCACCTGTCACTTAGCATACGCTGACAGCATCCAGCAATAAGCATTGCTGGGCA[A/T] GATTTCCCAAGGACTGGATCGCTTGAAGCTGCTTTACTTCTCATCTACAAGCCTGCAGA
S1_199164936	5	6	10295314	1E-37	ATTATTATTATTATTCTTCTTCTGACGTAACGGGTACATTGCTTGGAAAGAAAGTTG[T/C] TGGAGCTTGAAACCGCAATAATGATACGAACACAGCCTCAGCAATGCAGGATTACTCGCT
S1_282211588	5	6	17925427	2E-40	CACCTTGCACAAATATCTGCTGTAGAAATGTTCTATTGTTAAACCTGCAGATTAGGAAAT[C/T] CTAAATCTATCTGCTGTATGAAGTCTGGTAGTATGATCAAGCAGGTTGATTATACAT
S1_116006917	4	6	21845952	8E-52	TTGTCAAGGAACGATCCCTTCCACCTCTCGGAGAGAAATGCGCCGAACACACAGAGATG[G/C] ATATGGCCGGAACCTGCAGAGGAGAAAGCGAAGCCCTAACCTTGAAGTCTTTCAGCACAG
S1_45761963	5	6	27083504	1E-50	TTTTTATCATGAACCGATCATCTGTAGACAGGTAGAGAAAGCTCCCACTCTCCAGGGGA[G/A] GATAATTCCTCAAAGCATCACTTCCACAGATCGTCAACATAATAATCTGCAGCATCAACC
S1_289563297	3	6	31210760	2E-53	AATGAACCATATATACTAACAATAGATGTGAAAAAGAATAATTTGCAAACTGCAGGTGG[A/G] CAGGAAACCAAGGGCTAAATAGACACACCTCATGACCTAGTTCACACCCATCTCCTGT
S1_210284742	5	6	32022412	1E-50	AAAAGACCAAGGAAATGACACAGCAGAACCAATTTGCCATTGGACATTTCAACTACAT[T/G] CAAACCTGCAGCATAAAAACCAAGATTTATATACATATCCACACTAGTTCAATGAAACA
S1_244041680	5	7	891269	4E-49	ATCAAAACATCGCTCTGACGCCAAATCAGACGTTAGAGAAATATTATAGGCGAGT[G/A] ATGGCTTGTGTTCTAGATGGTACAAGATTGTGCAGCAAAGGCTTCGAGTCGTTTTG
S4_1831247	3	7	3180748	5E-48	GACAAACAGAAATCTTCCCTTCCATTAGGAGCAATCCACCAAGGAACTGCAGT[T/A] GCCAAATGAATCCGAGGGTCCCAACCACTGGCTTTCAGGATTGCCAAGCGA
S1_156520859	4	7	10367143	8E-52	TCTTTACTGATATAAGAGACTACCAGAACTCCATTTGTATGTTGGTTAATCTGCAGACA[C/T] TGAAACTCTATTGTTGTTATAAATCTTCCGAGCTTCCCAAGAGCAATAACATACATGAACA
S1_109907043	3	7	15658966	2E-53	AGGAGAAAAAATTCATGTGATGTCTCCATATCTCAGCCTCGTCTCGGGTGTCAAATGA[A/G] ACTGCAGCAAGTATGGACAATTCGAAAGATAGACATGGATGGCACTCTCAATGTGAGT
S1_365833705	3	7	17541018	1E-50	ACACATCTCACCATTC AATCACTTTTCATCCAACTGCAGCAACCTCAACAAGATCTCC[C/A] TGAGCTAGGTATCATCAATTTTCTACAAGCAATCTGCATTGGAAAGTGATCATGGACCGA
S1_215375978	5	7	17828247	4E-49	TGCTGTCTACGCCGATGGATACTGTAAGCAGCGGGCTGCAGCTTGAGAGTAGTCCGTA[C/T] AGAGGGTGGGTGATTGTGTGAGGAGAGTGTAGGGGAGAGGGGGTGCCTGCGGTTTTAT
S1_5956960	5	8	1990861	5E-35	AGATAAGCACTTGTATCTTGTCTTTTGTGCTTTTATTATTGATGCAACAATG[C/T] CCCCAACCCACACACACACACACAATTTTGTATTTTATGTTAGCTACTTTCAT
S1_142832546	5	8	4073006	4E-49	AACATAACATGAATTTGGTCAATGATATAAGATTAACAACAACAAAACGTTTTTCTGCAG[G/A] GTTCTTGAACAAGTTTGTATGAAATCACAAAATGGATATTGAAAGTTTGAAGAATGTTAT
S1_102926938	3	8	5771893	3E-50	AATCTTTGATGACAAAGCTGCAGCTCTTTTTCATGCAAAAACAATAAAAAGTATACCGGAT[C/T] TGATGTGATATGGGATGATCAGATCACTACTGAAAATGAAAACCTGTGCCAGCTTCTCT
S1_29231327	5	8	6476874	1E-48	TCCAGAAAATAGTGGGAAACATACAGGGCCAGGATGTTTCGAAAATGCACC[C/T] GCTAACATCGTGGCCAAAGCTATGGAAAACATTCGAAATGAAAAGTATAACCTGCAGCTGA
S3_54678463	5	8	7748403	4E-56	TCAAGAGCTTCAAGAAAGGAAAGGATAAAGTGAATATACAGTTAGAGTTAGAGTGTGG[G/A] AGACCTGCAGAAAGCAAAAGAGTTTGTGTACTGAAATGATGGATTGTGTTATGATCA

(Continues)

TABLE C1 (Continued)

SNP_ID	# Fluo.type	Chr.	Pos.	E-value	Sequence
S1_208236889	2	8	9479207	4E-49	AGGAAGGAAAGAGAAAGAAGTTTCTGTCGAGTCTCAGCCCTCTTCTGAATCTTCTTG[T/C] AGTTTATCAACAATCACTCAATCATCGGTGCAATGCCACTTCAAAATAACTCAGCAT
S1_169356495	5	8	12133164	8E-52	ATACAGAAATATCAGTGTATATATACAGAAAGTAGAATGCTTCATCACCGAATCTGAT[A] TTTATATGAAAACACACTGACCTCTTGATGAAGAATTAGGCAAAACAGGGAGTCTGCAGA
S3_35309770	5	8	19352521	2E-47	CATTTCCAAAATTTCCAGAAAATAAATCGGTTTCCATAAGATTGAGGTACAAAATGTTTCC[A/G] CAAAGGAGGTTTATCTGATACAACACTGCAGCTTGAATATGGTAATAACTAGTCTCACA
S1_235419648	4	8	23125222	5E-48	GAACTTCAGAAAATGTTATACGCTGCAGATTGCCAAAATGAAGCATTATAGACATAA[C/T] TGATCCCAATAATCAGCCAGCTTCTTTTATGTTGTACATAAAAAGTTCAATTAGCAAGAT
S2_30875426	3	8	27554786	1E-43	TTTCATGTTTGGAAAGATCAATGTCAATTTAGATGTCATATGTTTAGTTTTGTTATAGTA[C/G] TTTATGTTTGTATTATCCAAATAATCAATATCATAATCTGCAATTTCTGCAGCAGGTTCTT
S1_71285261	4	9	1039396	2E-53	AGAGAATGTCGGATAATCCTCAGAGAAGACTCCACCTTCGCAAGGTGCCAGCTCGCA[C/A] GGCCATGAAGAACCTCGAGCTCTGTGGGTTACGGCTGCAGCTCAGGCCATGCCCATCTT
S1_352413390	4	9	3168774	2E-53	CTTTCTGGCAACAGTCTGCAGTAGATTTGAAGTCAGTCTTCTTACAAAGTCTACCA[A/G] AGAGGATTCCAAGTCAGTGAATGGATATGATCAATTTGCATGACTGCTTGA AAAAGTCGGG
S1_58454213	5	9	3570970	2E-53	TGCTCTGGCCTCATCGAGGAGCCATTTCTAGCATTTGATAGAGGAGGTTTATGG[C/T] TCTCAACTCTCCTTCCCTTCAGAATCACCATTGAAGATTGCTGATCTGCAGATGATT
S2_9856110	5	9	5066064	1E-50	TGGAAAATTAGGTATCCAGTTACCATTGAAAATCGTAGTGTGCTTGATAGGCAAGG[T/C] CCAATTTACAGAGAGGATACTGCCGTGTTCTGTAGCCAACTGGAGAACTGCAGATACC
S1_157448006	5	9	7773168	2E-46	TTTAACTTTTGAAGGCTGCAGGATGAAATCACAGGCCCGAGCTGCTAAATGTAGAT[C/T] ATGGTGAAGAAGCTGCATCTGAGGATGAAGAAATCTTATGATGGACATGATGCAGATG
S3_14699960	5	9	14771343	1E-43	AACTAAAAGCAAACAAAATAAATTCCTGCGGTTAAATAACCTGCAGAAAATAAG[C/A] GAAATGTGACAAGGAAATATTACATACTTCCCTCCATGACACTTCTTTCAGAGTTT
S2_58843160	5	9	17855377	5E-54	CAATTCATTTCAAGGCATAATGTTATCAAGTAATGCATATTCTACCAGAAATGAACTTTAT[G/A] TGGAAATCATCTTGACATTTTGAAGAAGTGCAGTTGATTACAAGTGAATGCTTATAAC
S1_108505610	5	9	23138342	1E-48	GGAAAATATCCAAAAGCAATAAAAAGCTGCAATGGACGCGCAATGCCCTGTTTCCACA[A/G] TCGAAGGCATTCGCCTAAGCCGCTCGATGTTGGTCTAGACACCACCTGCAGTTTCGAGAA
S1_96409136	4	10	273696	9E-45	TGAAGTTTGTGATTCATTTACCAATGATTTTCATGCACAGGATATATCAGAAATA[C/T] GAGCAAAAATGATGCAGAGGTTACTTCAGCAAGAAATGCTGCAGAGAAGGATGCACCAAG
S1_75907479	5	10	1295979	8E-52	AATGCTTGACAGAAAGCCATGAAAAGCTCCAAAAATAAGTCCAAATGAGATTGGAGA[G/A] CATACTCCACTGCAGCAAGTGTACCCCTGTCTATGTACTGCTGGAGGGGCTTTGC
S3_17911820	3	10	2220748	8E-39	AGTAGTTTCTGAACTGGTACTTTGATCAATACCTGCAGAGTTAGTAGCAGTAAAG[C/T] GAGGAGACCTCAATATCTTGCACATGATCACTCACCGAAAATGGAACATATCAGCATGA
S1_282032037	4	10	5002351	7E-40	GATTTACAGGACAATACATTTACAGATTTCCATAATGATGTTAACTACAAGAAATATT[C/A] TGTAAGCACCATGATACTTGATCCATTACATGCAATGAAATCAAAAAGAACTGCAGTTTTA
S2_66969333	2	10	5976818	2E-52	TTTAACTGTATTGGCAGTGTCTGCAGACAGAGCTACGCTAGCAGAAAGTGCAGCAACTAT[C/A] TCTGAAAACAACCTCCAATCTGTAAAACAGAGCAAGTCAAGAAAAATTTATACCAAGATAAG

(Continues)

TABLE C1 (Continued)

SNP_ID	# Fluo.type	Chr.	Pos.	E-value	Sequence
S2_53836832	3	10	10477461	5E-48	GAACACTTCTCGGATGAAAATATTCTGCTGCAGATCATCGTTCTTTGGCACCACA[C/A] CCTTTTCTAATAAATATCTTGAGCTCTTCTCATCTTTGAGATGAAAACATCTAAC
S1_21668183	4	10	16038685	2E-53	AGTCGAGAAAATACATCAAGATGGATCTCAGAGCTTCCAACCTGCATAACATCTCATC[G/A] ATTGCAATGTCATCAACTGGAAGGTGCCCTATGATCCCTGCCACGGTTGATCTCCTCT
S1_333790152	5	11	2622759	4E-56	TGCTTTGAAAGGCCAGCATGATCTTTTCTATTTTGTTGTTCTTGAATGAGTTCTGTT[A/T] ACATTTGCTGATTTTGTCTTCTTGCGAGTCTGCAGTAGATTTTGATGATAAAAACAGTTGG
S1_238131512	2	11	14447426	1E-43	AAATATCGAGATGAATTTCTGGAAACAAGCCTGCAGTTAGTTTGAGGATGTTTGGAAAT[A/T] AGTGATCAAAATGGCATTTGAGAAGCATAGTAGTAATTTTCCGTAATAAATGATCAAGA
S1_38918393	3	12	17162245	2E-46	AATCCCATCAAAATTTGCTTTCAAAATTTCTGAATTTTCCCTATCCCAAAAATTCAG[C/G] CCATCAAGAAAATATAGGCAAAAGCATAAAAACCTGCAGCATAATCTATTAGATCTCATC
S1_102041015	4	12	24327364	3E-45	GCAGTATACAACCTTCAATACCATGTTAGACCAGCAAATCTCAGTCTGATTTCTTCCTAGC[T/C] AACACACACACACACACACAAAAAGAAAACCTTCAATCTCTGTTGTTTCTGAGTGCAT
S1_65460849	4	13	1824700	1E-42	GCATCTGCTAGTTCAAAAATAAAGGCAAGCATGGTATCACTAATACTGCAGAAAAATGAT[A/G] GTGCATAAATATCTAAATGGGAAATGATGCAGCGAAGATCAATAACTTAGAATGAAATTC
S2_23695000	5	13	8252985	8E-52	TACATCTGGGGACTGCAGAGCTTTGAGCATCCACTGAACCTGGTGAACCAGTCCCAG[G/A] GTTGACAGCAAAGGGCAAGTATGGTGCCAATTTCAAAGTTGATGCCAAAGCTAAGAAG
S1_279910017	5	13	27313051	2E-53	GGCAGCTTCCAGGGCTCGGAGCTACCAAGATGGAGTTTGGTATCCACGGATTTGTTCCA[C/T] GAATCTGAGCTCCATCGTATACCCTGAGCCTGCAGCGTCTCTGCAGTCCAGAGCATAA
S1_297529258	3	14	3933202	7E-46	CTTATCTATGGCCAGGTTTCAACAATATAAGGCTTCAATGGTCTCAAATTTATCGGGT[A/C] TGCTTCTGCTGTGTTACTTTTACTTGTACTTACCTAGCCCAAAGTAACTGCAGCCTGCAGGT
S1_252699764	5	14	12391182	4E-49	ACTACTGAATAAATGGAATCAACTTTATTTGCTGCAGTTGGACTAGCCTAAGAGAACTA[A/T] GTGGCTTTGGAAGAGTGTGATACTTGGGATTTTATATCAATGTGAAAATCAGTGACT
S1_64347285	5	14	12868762	1E-50	GGATCAAAAGTTCTCAGAGATTAATGATTTTGAGAAGTCAAGATATGCATAAATCCGTGG[T/G] TGATCTCTACTGCAGCATTTACCATCCCTTTCATGTCAATTGACCAGAGAGGTGCAGAT
S1_108652759	4	14	13736821	3E-45	GAGGTAATCTGTGACTTGTCCATATTACTGCAGAACAGCAATTTATTTGCTGATCTTGGAC[T/C] ACTGATATCCAGCTTTCCCCAGATTTATGTCATATTGACCCAGCAAACCCAGTACAATTTA
S2_68074878	5	15	320619	6E-47	CACCACCATCACATCCGCACCATTTGTTGCCACCTTTTGAACCCAGTGTGCCCTGAGAAGGA[C/T] ACCTGCAACATATCCGGTGTGATGACTTCTGGAAAAGTTGCCCTGCAGGGCAGATAATAAAAAT
S2_15031349	4	15	2754888	3E-50	TAAGAAAAAATAGCTTACATTATCTGTGTCGACCTCTGATATAATCTCTTTGATTTAGT[T/C] GGCATTGCCCCATTCGGTATTTCTCCATGGCTGATTTCCAACCTCATCTCTTGATATAGCT
S1_361168697	5	15	3199245	2E-54	ATTACTAAGATGCAATTTATAAACATATCTGCAGTTTCAATTTGAAATCTCTGCACA[A/G] GCAGAGAAGTTGAGATTTCCGTAAAAAAGCCTGAAAATGGTGGTAGTCTTCAGAAGAG
S1_42812024	5	15	3456280	8E-52	GAACATTTATAGTACCTCAAGGGGAGTGCCTTTACTGCAAGAGGAAAGGGAACACCCA[C/T] TCCGCATCTCTCGGAAATCTGAAGAGCTAGGCCCTGCATCAATGGCTGCTGCAGTAGC
S1_31917523	5	15	3839679	2E-54	TTTACATCTAATGAACCTCTGCAGGTTGAATGAAATATTTGCTTGCATGGTGGTTT[G/A] TCCCTTCTAATGAGACCTTTGATAACATACGCCAATTTTGATCGTGTTCAGAAGGTTCTCT
S1_16064479	5	15	5173749	1E-50	GCTCAGAAGAGCTCCATATGAAAGTTGGTTTCTGGACTACCTCCGGGAGACTATTGAAGTA[A/T] CTCTCTGCAGCATACCCCTTTGACTTTGATATAAGATCTATTTCGTTATTCATGGGTC

(Continues)

TABLE C1 (Continued)

SNP_ID	# Fluo.type	Chr.	Pos.	E-value	Sequence
S1_116148629	4	15	6642827	8E-52	GCAGGAATACAAGAGTATTGCCAGAATGAGGCTGGTACATATTAGCCAAATCCGAATCC(A/G) AGTCATTTTACAGTCTTGAGCTTCAATTCAAAATCACCTGAGGAATCATACAGTGAT
S1_359692995	4	15	8357684	4E-49	GTTGGCTGCTAATTTGCAGTACTGCAGAAGTAAATATGAAAAAACAATGAAATGATAATCA(C/T) AATTCATTGACTAACCTGTGCAGATCTAGATAGTAGATGAGGAGCGATTCCATTCATC
S1_290181714	4	15	9566725	2E-54	TCCTCGTATGGGCCCCAAGGGAACCTCAAGTTTGCTGGAATCCTTTTGGGATGGGA(G/A) AGCAGTGTCTGAAGATCGAGAAGTTGCTGCAGATCTCAGGCTCCATTTGGAAGCAA
S1_282323032	4	16	4930222	4E-49	ACTTGGGATTTGATGATCAGTGTGCTACTATTTTGTCCTTATGTCCTTTGTCAAAAGAA(C/T) CGTAATAGTTCTGATTCAAAACAATAAACAACCTGCAGGTTGCTTTCATTTGGATTTGA
S1_262914420	4	16	8184285	4E-49	AGCAGCACTGCAGCAAGAAAATAAACAACCTGCAGCAGAGAAATAAACAACATCACCA(T/A) GAGACTTCAAAAACAACAATTTAGTCAACAAGGTTCAAGAAAACAAGAACACCTTGCA
S2_63978772	5	16	14079687	2E-53	AAATACTGGTCAATAAATAATAGATCACACATATGTCAGTGTGGTCAACAAGATAATGCG(C/T) TGCTGCAGTAAAAAGAAAGAACTTCAGCAGTGTGGCTAACAGACCTATTGTAACACTAGCA
S2_42975314	5	16	19099744	5E-54	CTGGGCCACGAGGGGGGCTGGTGAATAAAGCTGACTGGAATAAAGCTCCCTTCACAGCCCT(G/A) TACCGAAACTTCACTGCTGATGCTTGTGTTTGGTCACTGGCGCTCAAGCTGCAGATCA
S1_349651036	4	16	22229815	5E-54	GCSTTGGAAAGATAAAAATTCATCGAAAGTCAGAGAAGAAAGCAAAATTACACAATTCGTC(T/A) GCTGCAGTTCGTCCTTAGTTCATCCACAGGCCCTTGGATATGACAACTCCCTTCCAGGA
S1_131821504	4	17	11148474	6E-41	ATCACTAGTACAATGCAACATGACAAAACCTGAAAGTGTATACAAAGTACTGCTTATTT(A/G) AACAGGACAAATCTGAAGCATAGCTGTAATACTTTCGCAGAAAAATAATGGCAAAAGTTT
S1_305511589	5	17	13065352	5E-48	TGATGTAGATAGTGGAACTGATTTTATAACTTTATTACAAAAGCATATTTT(A/T/G) CCATCTAATGTTCTGCAGATTAGTGGATGCTATTTTATAAATTTTTTAGCAATTT
S1_162377692	4	17	14256863	2E-53	ACAGAAGAAAAGCTTGATGAGATGATGATCAGTTGAGAAATGAGTATGATGAGTAA(G/A) CGATCAGCTATACAACCTGCAGGCAACTTCTCCAAAAGAGCTGACCCAGACTTGTTTTCA
S1_7095019	5	17	18481610	3E-44	CACGGGTGAGCCATCAGTAGTACTGTAGCCAAATGTTGAAACCATCGAATCCCTTTCAC(C/T) GACTGCAGGTATACGACCTATAGAAAGTTGCGGAATTAGTGACTACTTTTTTTTTTTTGT
S1_164752250	5	17	19197258	1E-50	CGTTCAACAGCAAACTGCAGAGAACATCAATCAGATGATCGGAGCGAGCTCATCAGCAGA(T/G) AAAGCAATTGATGATGTTTTCAGTTTTGATCCAAAGCATCAGTAGGACCTGTCCAG
S1_104430414	4	18	7027601	6E-47	GGTACACATGCAGACCTGGAGTCTTTGTTCCCTCTTATAACATGAGAGCTTGTCTTCT(T/G) GTCATATACTTGCTGAAGCTTTTGCAACTTTTTTACATCTGAAGCCAAATTTCTTTACCC
S2_23048737	4	18	14902976	8E-52	TGATTATGCTCGCCATTTCTTGCTGCTGCAGTAGAATGCTTTGGCCTGTCTTATGAATC(C/A) AAGAGAGCTACATTTGGTTTGGAGTACTATGGCAGGACAGTGCAGTCAAGATTTTGCCT
S1_259660238	3	18	18609390	8E-52	TTGTAGGCATCAATCTCCAAGAAATCATATCTATTGATGTTGATGTTGTAAGTCTTGTGAGG(C/T) TGATAACTGCTGCAGCCCTCTCTTAGTTGAAAGCATAAACCTTTTTTCTAATTTTCCCTTTT
S1_170367846	5	18	21499070	8E-52	CCATGCTCGGCCCTGCAAAACATTGGATACCAAGAAATATAAGAAATGAAAGTGAACGGAT(G/A) CGATGGTGA AAAACAATTCAGTACCTGAAAACCTGCAGTAAAGCACTTCTGAAATGAGTTC
S1_31156518	5	19	1787283	2E-53	AGGACGTCACACACTTGATAGTGTTCATTTGTCATGCTTTCATCCACATGAATGCCT(C/C) GGAGACATGATGTTTTCAGTGTGCCAACCTGCAGTTTAGTAAACATGAAATGACATACTACAA
S4_1766101	5	19	2912318	2E-41	TTCAGATGGATCAGTCGTAGATTGAGCTTTGAAATCTGAAACAGGAGCATGACCGATAG(G/T) AAGCGGAAAGCAGGAACAGGAGAAGGGAAACATGATCTGCAGCTGCAGCCACCATATAAGA

(Continues)

TABLE C1 (Continued)

SNP_ID	# Fluo.type	Chr.	Pos.	E-value	Sequence
S1_130386685	5	19	3858772	2E-54	AGAGTTGATTATAAATTTATGATCACATAATAATTGAACAAAAAAGTGCAGAAAAACAATACC[A/G] GCTAAGTCATTGAGCACCACAAATATTAGAGGTAAGTTTGATTACCTTCAACTTTCAAG
S1_66878388	4	19	4662624	6E-47	AATTGTGACATCAGCAGCAATAAAAAATTTTGAACAAACCCGCCCCAAAAAATTTATCAAAA[A/G] AAAAATTATCACTGTCTTGTGTCAAAACATCATCCAACTCCAAACCCATACTGCAGAAAA
S1_48170206	4	19	8029141	3E-50	ATTATACAAAAGTGCACAAACGTTATTCATATCAAAAATTTGAATTTACAGGAGAGATTGGTAGG[A/T] CGTAAAGATCCAAACATAGAAAATCAAGATGTTAGAAGGATATGATAGAAAACAACATGCAGA
S1_85054393	4	19	9504285	1E-50	CACATCTGTTTTCTCTTGGAGATTCTACAATAACAACCTGAGAAGTTCTGCCAAAAAATTT[G/A] TCTCTTTGCGGGTAAAATATTTTCTCTCCATAGGAGTGACAACAATAATCTGCAGGCTT
S1_16428287	5	20	725184	8E-52	CAACGTACTGCAGACCCGTCAAAAATGCCAATGCAATAGGTTATTATAGTTATGAACATCTC[G/A] TCAACAAAAGATTTGCAACCAATGGTATGAGCATTTCAGATTTCAAGACTGTCCAAAAAACC
S1_43488339	5	20	2904117	1E-50	GACAAAGTCATCAAGTAATTCCTTCATGACAGATAAATCTGACTGCAGGGATGTGCC[G/A] AGCTTATAAGGCCAATTAGACAGCATAAAACAATAAGAAATGCTGGTCAATGAGACAATG
S1_114295977	4	20	6774493	1E-48	GCTGGCATGTTATGATGAGTATTCCTTTTGATAGCAGTAAAAATAGTACTTATGGCT[A/G] ACTTGAAGATCTAGAAGATTGGTTGGACTGCAGTTATATTTTTCTTAGCTCAGCTAGG
S1_30976509	3	20	8500327	2E-54	ATTGTTGGCTCCATCAATATAGGAAGATATCTAAGGAGTTACTGACTTGTAAATGGA[C/T] CATAACTGACAATAGCTTCAACATGATGACTTACTTCGTGAATGCCTGCAGAGGAGG
S5_17384653	5	20	8742611	1E-36	TAAACATTAATAATTTTACAAATGTAATAATTTCCACAGAGAGTAGGTTCTGTAATCTC[A/A] GATTTCTTCGACAAAATCAGCGGATTCATTTTATCTGTTTATTGTGCTCGCAGCT
S1_160073245	4	20	9514825	1E-42	TTCCGCTGATGCTTAAACATATAAAAAGCTCAGATGTCAGATGCAAAAATAAATATCGACATG[A/G] TCAATAGTTAGATTAGCTTTTACGCTTGTAGATTAACAATACAGACCCACCAATAAAG
S2_46633476	3	20	9808073	1E-50	AGATATGTCGTGAAAATACCTTCAGATCCTTCTCCTCCCTCCCGCTTCCCTCGGCTCC[A/A] CCAAAATTCGCCCTTTCCTTCCACCCACCTCCGCTCCGCTTATGATGAGCAGCAGTGA
S1_231563096	5	20	9944570	4E-49	GTTGAGTTTCAGAGTTCTGCAGCTCCCTTTGCAGTAGTTAGGCAGTAAACCCATCACCTG[A/G] AAAGCATATCGGTTTCGATACTGTATGAACCTCCACAAGAACAAGCCCTGCCATTGTCTCCA
S1_60794376	5	20	10523563	7E-46	TTCAATAC TTGTTGTCATTATAACATCAGAGATTGCAATTCCTTCATGATGCCACTGCAG[G/C] ATCCCCATATGGACAGCTGGTTTCAGTTGTCATTATCCTTATGCACTCAAATCTTGAAAT
S1_199666449	5	21	240474	4E-49	AGTTCACATCAAAAACCAGTTCTCTACCCAAAGGGTAGCAGAAATTCCTCAACAAAAT[A/G] TAGAAATTCACATTTGTTCTCCGGAAGCATTGCCATTGGTGGTGCAGTTACTCCATA
S1_257000706	2	21	928527	2E-54	TACAAAATGAGATGGTTACAAAGCTGAAGAAACAAACCTCATCTGCAGTAAAAAGAAAGAC[A/G] TCACTGTGCCTCAAAAAAAGTTCAAGAGCTCTCTGTTGTGAGGCAGCCCAAGCAAAAAGA
S1_171090323	5	21	2104297	1E-50	CGAGGAAGGAGAAATGGAGCTACTGATTGCAGATACTTATTGGTTTCAGCAGACTAC[G/A] TCCATCATGACCCCTTATAATACATTTCTGCAGGCAATACAGTTCTTCTTAGGACC
S1_25663258	3	21	2772381	8E-33	TGTGAAAATTTGGCTGTTTGTCTGCAGTTTCATCATATTTGCAAACTAATAATATAAACTC[A/T] AAAAATTTGTCCACTATAGAACTAGTGCAGCATCAATTTCCAATATAGCCAGGCTAA
S1_371159620	4	21	3932249	1E-50	GTCTGACGGGGAGGTTCAAACTGCAGCAGGTTTCGCATCAATGAGTGTAGCACCTTTATCA[C/T] CAGCATTATTCCGGCTGTGCCGCCACCACCTCATGTTGGACTGCATCCAAACAATCCTGA