Quantitative trait loci and candidate genes for physico-chemical traits related to tuber quality in greater yam (Dioscorea alata L.)

Gemma Arnau, Lucienne Desfontaines, Adou Emmanuel Ehounou, Carine Marie-Magdeleine, Amani Michel Kouakou, Jocelyne Leinster, Elie Nudol, Erick Maledon, Hana Chair

Research Article

Received: 18 January 2023 Revised: 16 May 2023 Accepted article published: 4 July 2023 Published online in Wiley Online Library:

(wileyonlinelibrary.com) DOI 10.1002/jsfa.12822

Quantitative trait loci and candidate genes for physico-chemical traits related to tuber quality in greater yam (*Dioscorea alata* L.)

Gemma Arnau,[a,b*] Lucienne Desfontaines,[c] Adou Emmanuel Ehounou,[d] Carine Marie-Magdeleine,[e] Amani Michel Kouakou,[d] Jocelyne Leinster,[c] Elie Nudol,[b,f] Erick Maledon[b,f] and Hana Chair[a,b]

Abstract

BACKGROUND: Starch, dry matter content (DMC), proteins, and sugars are among the major influences on yam tuber quality. Genetic improvement programs need simple, rapid, and low-cost tools to screen large populations. The objectives of this work were, using a quantitative trait loci mapping approach (QTL) on two diploid full-sib segregating populations, (i) to acquire knowledge about the genetic control of these traits; (ii) to identify markers linked to the genomic regions controlling each trait, which are useful for marker-assisted selection (MAS); (iii) to validate the QTLs on a diversity panel; and (iv) to identify candidate genes from the validated QTLs.

RESULTS: Heritability for all traits was moderately high to high. Significant correlations were observed between traits. A total of 25 QTLs were identified, including six for DMC, six for sugars, six for proteins, and seven for starch. The phenotypic variance explained by individual QTLs ranged from 14.3% to 28.6%. The majority of QTLs were validated on a diversity panel, showing that they are not specific to the genetic background of the progenitors. The approximate physical location of validated QTLs allowed the identification of candidate genes for all studied traits. Those detected for starch content were mainly enzymes involved in starch and sucrose metabolism, whereas those detected for sugars were mainly involved in respiration and glycolysis.

CONCLUSION: The validated QTLs will be useful for breeding programs using MAS to improve the quality of yam tubers. The putative genes should be useful in providing a better understanding of the physiological and molecular basis of these important tuber quality traits.

Keywords: dry matter content; starch biosynthesis; molecular breeding; heritability

INTRODUCTION

Yams (*Dioscorea* spp.) are herbaceous vines cultivated for their starchy tubers. They represent a major food crop in the tropical and subtropical regions of Oceania, Asia, the Caribbean, South America, and in particular in West Africa, which accounts for 92% of the world's production.[1] *Dioscorea rotundata* and *D. alata* are the most important cultivated species. *Dioscorea alata* ranks second in production importance, but it is the world's most widely distributed species. This is due to the fact that it offers several particular advantages in terms of early vigor for weed control, yield potential under low to medium soil fertility conditions, and a better tuber storage ability.[2,3]

In West Africa, yams are consumed in several ways (boiled, fried, roasted, baked, or pounded, after being cooked, into a stiff paste called *fufu*). Cultivars of *D. rotundata* are more suitable for the preparation of *fufu*. However one study showed that tubers from

* *Correspondence to: G Arnau, CIRAD Montpellier Centre, UMR Agap Institut, F-34398 Montpellier, France. E-mail: gemma.arnau@cirad.fr*

a *CIRAD, UMR AGAP Institut, Montpellier, France*

b *UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France*

c *INRAE, UR 1321 ASTRO Agrosystèmes tropicaux. Centre de recherche Antilles-Guyane, Petit-Bourg, France*

d *CNRA, Station de Recherche sur les Cultures Vivrières (SRCV), Bouaké, Côte d'Ivoire*

e *INRAE, ASSET, Petit-Bourg, France*

f *CIRAD, UMR AGAP Institut, Petit Bourg, France*

some *D. alata* genotypes could form a good dough, comparable to that of some *D. rotundata* genotypes.[4]

The sensory quality of tubers depends on many physico-chemical and textural characteristics.[5,6] It has been shown that the quality of tubers for consumption (either boiled or pounded) was strongly related to their dry matter content (DMC), starch, and amylopectin content.[7] The four main constituents of tubers are starch, DMC, sugars, and proteins. Tubers of *D. alata* contain 20–40% of DMC, 60–80% starch, 0.5–11.6% sugars, and 4.0–7.4% proteins.[8]

Genetic improvement programs need simple, rapid, and low-cost tools to screen large populations. Near-infrared reflectance spectroscopy (NIRS) has been demonstrated to be a reliable technique to predict the major tuber constituents in *D. alata* yam species.[9,10] However, as spectra were generated from flour samples and not from raw samples, this protocol requires a long sample-processing time and remains rather difficult to apply to a large number of genotypes.

Marker-assisted selection could be a high-throughput method facilitating breeding efforts. Indeed, with the development of new-generation sequencing technologies, it has become much easier to search for genomic regions associated with traits of interest. A few studies have been conducted with yam to elucidate the genetic determinism of tuber quality related traits. By using a quantitative trait loci (QTL) mapping approach on two biparental populations, several genomic regions linked to important morphological and agronomic tuber quality traits have been identified.[11] The heritability of DMC was estimated on a diversity panel including eight different *Dioscorea* species.[12] A Genome-Wide-Association study was carried out in *D. alata* and some single-nucleotide polymorphism (SNP) markers linked to DMC could be identified.[13]

The current study used a QTL mapping approach on two diploid full-sib segregating populations, aiming: (i) to acquire knowledge about the genetic control of starch, DMC, sugar, and protein content in tubers; (ii) to identify markers linked to the genomic regions controlling each trait, useful for marker-assisted selection (MAS); (iii) to validate the QTLs on a diversity panel; and (iv) to identify candidate genes from the validated QTLs.

MATERIALS AND METHODS

Plant material and sample preparation

Two diploid full-sib segregating *D. alata* populations composed of 93 (population A, 74F x Kabusa) and 80 (population B, 74F x 14 M) progenies, respectively, were used to map the quality traits. Both populations were generated previously[11] and derived from crosses between diploid progenitors contrasting for expression of quality traits. The male Kabusa and 14 M parents have significantly higher DMC and starch content than the female parent 74F, which has, on the other hand, higher sugar content than 14 M and lower sugar content than Kabusa. Progenitors do not differ significantly in protein content.

The mapping populations and progenitors were planted in two blocks at Roujol experimental station in Guadeloupe, France (16° 10′ 56″ N, 61° 35′ 24″ W, 10 m.a.s.l.) during cropping season 2017–2018. Each block included nine plants of each genotype. After harvest, about 200 g of the central part of three to four tubers per genotype was sliced into chips and dried in an oven at 60 °C for 72 h (3 to 4 tubers × 200 g × 2 blocks). Then flours of each tuber were prepared for NIRS screening as described in Ehounou *et al.* (2021).[10]

The diversity panel used for the validation of QTLs consisted of 24 *D. alata* genotypes and included landraces and breeding lines presenting a high diversity for the four traits whose qualities were studied. These accessions were planted together in the same field during two cropping seasons (2016–2017 and 2017–2018) in Guadeloupe (16° 10′ 56″ N, 61° 35′ 24″ W, 10 m.a.s.l.). After harvest, as for biparental populations, about 200 g of the central part of three to five tubers of each genotype were sliced into chips and oven dried at 60 °C for 72 h (3 to 5 tubers × 200 g × 2 years). Then flour from each tuber was prepared for chemical analysis as described in Ehounou *et al.* (2021).[10]

Phenotyping of progenies and the diversity panel

Phenotyping of 24 accessions of the diversity panel was carried out using the chemical analysis and methods described in Ehounou *et al.* (2021).[10] Starch, sugar, and protein content was expressed as a percentage of dry weight.

Dry matter content, starch, sugar, and protein content of progenies and progenitors were predicted by NIRS analysis, using a FOSS-NIRSystems model 6500 scanning monochromator (FOSS-NIRSystems, Silver Spring, MD, USA) and partial least squares models developed previously.[10] Two replicates were scanned for each flour sample. A total of 1762 NIRS measures were carried out in the technical platform of INRAE's UR143 ASSET research unit, in Guadeloupe. The NIRS spectra were generated from a single block for population B.

Statistical analysis

Pearson correlation tests, histograms, box plots, ANOVA, and normality analysis were performed using XLSTAT version 19.03.44616. Distributions of progeny phenotypic data were tested for normality using Shapiro–Wilk and JarqueBera tests. Broad-sense heritability was estimated as described in Ehounou *et al.* (2022).[11]

Genotyping by sequencing

Single-nucleotide polymorphism genotyping data of both mapping populations, progenitors and the diversity panel were generated previously.[11,14]

Quantitative trait loci mapping

Quantitative trait loci analysis was performed for each population separately using Map QTLversion 6.[15] Quantitative trait loci were detected using the interval mapping (IM) approach, mean values and the previously published reference genetic map generated from two populations.[16] Significance LOD score thresholds were calculated through permutation of 1000 iterations with an alpha risk of 0.05 and confidence limit of 95%. Confidence intervals of QTL positions were determined as two-LOD support intervals.

Quantitative trait loci validation

Based on the genotypic and phenotypic data from the diversity panel, QTL validation was performed in two steps. First, a simple linear model that associates each phenotype with a SNP was tested using the Pearson correlation test and a 5% significance level. All markers included in the QTL confidence intervals were tested. The Pearson coefficient of determination (R^2) and the *P*-value (Pearson) of each SNP marker were thus determined. Second, an ANOVA was performed for the significant markers to determine whether the observed genotypic classes were significantly different at $P < 0.05$.

Identification of candidate genes

Candidate genes for the validated QTLs were identified by searching near the significant SNPs in the NCBI database, which contains all 35 078 genes that were annotated on the *D. rotundata* reference genome.[17]

RESULTS

Phenotyping of mapping populations

The mean values and ranges of scores for starch, DMC, sugar, and protein content in the two mapping populations are presented in Table 1. Starch content ranged from 74.0% to 83.6% in population A (74F × Kabusa) and from 71.1% to 85.6% in population B (74F × 14 M). For DMC, the range of scores was from 26.7% to 38.8% in population A and from 27.6% to 41.3% in population B. Sugar content ranged from 0.58% to 5.08% in population A and from 0.10% to 5.84% in population B. Finally, protein content scores ranged from 3.95% to 7.64% and from 4.05% to 7.90% in populations A and B, respectively. In male progenitors, 14 M and Kabusa, the values were significantly different from those of female 74F for starch (80.5%, 79.2%, and 77.2%, respectively), DMC (31.5%, 28.9%, and 28.1%), and sugars (1.79%, 4.12%, and 3.22%).

The ANOVA on phenotypic data of population A showed highly significant ($P < 0.0001$) genotype effects and significant repetition effects ($P < 0.05$) for all traits (Table 1). However, genotypes were the most important source of variation for each studied trait. An ANOVA of phenotypic data from population B also showed highly significant genotypic effects ($P < 0.0001$) for all studied traits (Table 1).

The frequency distribution of all traits studied showed typical quantitative variation in both mapping populations and all traits fitted a normal distribution (Fig. 1). Transgressive segregations were observed with lower or higher phenotypic values than those of the parents for all traits (Fig. 1).

Broad heritability for the four traits ranged from 0.68 to 0.88 in population B and from 0.69 to 0.81 in population A (Table 1). The heritability for starch was similar in both populations (0.68 and 0.69) and also for protein (0.78 and 0.81). The heritability obtained for DMC and sugar content was significantly higher in population B (0.86 and 0.88) than in population A (0.75 and 0.72).

Several significant correlations were detected between traits (Table 2). Negative correlations were found in both biparental populations between starch content and protein content, and between starch content and sugar content. A positive correlation was detected in population A between starch content and DMC, while in population B, a positive correlation was detected between DMC and sugars.

Quantitative trait loci detection

A total of 25 QTLs were identified for the four studied traits in both mapping populations. For starch content, four QTLs were detected in the population B located on chromosomes 2, 5, 6, and 10, which explained 22.4, 22.6, 20.1, and 22.1% of phenotypic variance, respectively (Table 3). In population A, three QTLs were detected on chromosomes 10, 11, and 18, which explained 15.4, 19.1, and 16.1%, respectively of total phenotypic variance. Both QTLs identified on chromosome 10 were located in distinct regions and are two different loci. Figure 2(a) shows the one that was identified in population B.

For DMC, four QTLs were identified in population B located on chromosomes 1, 4, 7, and 12, which explained 98.5% of total phenotypic variance. In population A, two QTLs were found on chromosomes 1 and 2, explaining 14.3% and 18.7% of phenotypic variance, respectively. Figure 2 shows information from the QTL identified on chromosome 2 (population A).

Three QTLs were detected for sugars in population B (chromosomes 7, 9, and 13) and three in population A (chromosomes 6, 7 and 12), explaining 68.5% and 58.8% of total phenotypic variance in each population, respectively. Figure 2 shows results from the QTL found in chromosome 7 (population A).

Finally, four QTLs were revealed for proteins in population B (chromosomes 2, 5, 8, and 19) and two in population A (chromosomes 10 and 18), which explained 98.4% and 38.4% of each population's total phenotypic variance. Figure 2 shows the results from the QTL found on chromosome 19.

Quantitative trait loci validation

The majority of QTLs (22 of 25) were validated in the diversity panel. Table 3 presents the SNP markers located within confidence intervals of the QTLs, showing a significant association with the diversity panel phenotypic data (at $P < 0.01$** or $P < 0.05$*). The alleles at each locus and the allelic effects are also presented in Table 3. The analyses of variance showed that differences between the different genotypic classes were significant ($P < 0.05$) for all validated QTLs. Figure 2(b) presents the phenotypic data distributions of the different genotypic classes for QTLs

Table 1. Phenotypic variation for starch, DMC, sugar and protein content in mapping populations A (74F × Kabusa) and B (74F × 14 M)

Trait	Pop	Mean	Min.	Max.	G	R	H^{2a}
Starch	B	78.7	71.1	85.6	***	-	0.68
DMC		33.3	27.6	41.3	***	-	0.86
Sugars		3.03	0.10	5.84	***	-	0.88
Proteins		5.74	4.05	7.90	***	-	0.78
Starch	A	79.8	74.0	83.6	***	*	0.69
DMC		32.7	26.7	38.8	***	*	0.75
Sugars		236	0.58	5.09	***	**	0.72
Proteins		5.55	3.95	7.64	***	*	0.81

Note: ***Significant at $P < 0.0001$, ** $P < 00.1$, * $P < 00.5$ for the effects of genotype (G) and repetition (R) on the phenotypic variance estimated by ANOVA.
[a] Broad –sense heritability. Starch, proteins and sugars are expressed as a percentage of dry weight, and dry matter content is expressed as a percentage of fresh weight.

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

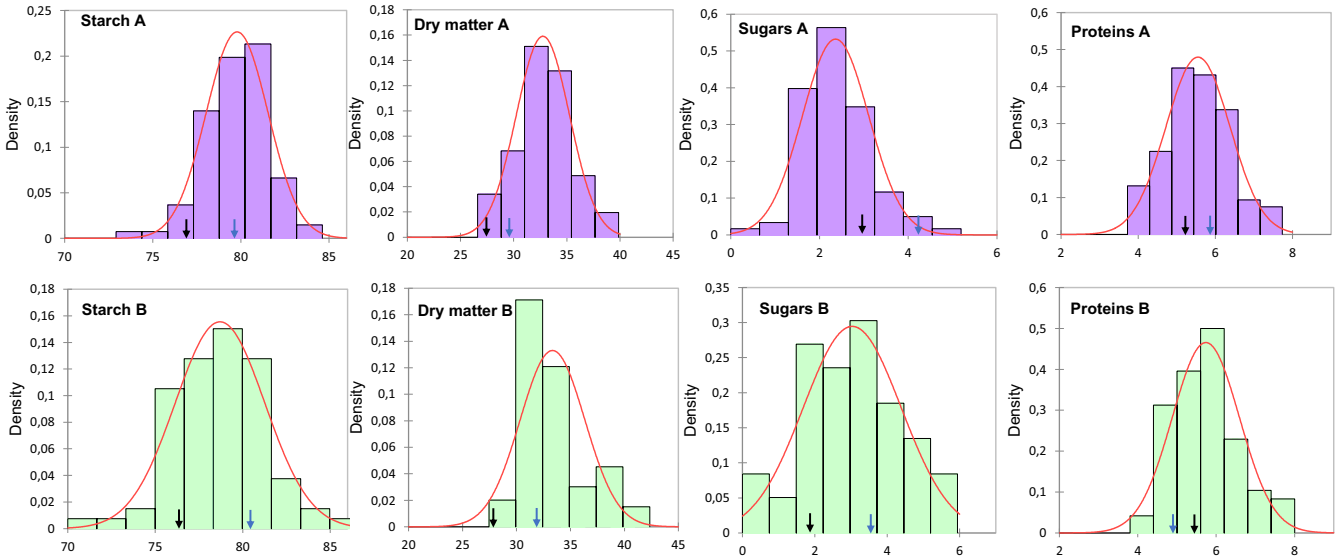

Figure 1. Phenotypic distributions for starch, dry matter content, sugar and protein content in mapping populations A (74F × Kabusa) and B (74F × 14 M).

Table 2. Coefficients of correlation between traits in biparental populations A (74F × Kabusa) and B (74F × 14 M)

Variable	Starch	Sugars	Proteins
Sugars	−0.34*		
	−0.57*		
Proteins	−0.66*	−0.11	
	−0.27*	−0.24	
DMC	0.22*	−0.15	−0.20
	−0.23	0.47*	0.00

Note: Top values are those from population A and bottom values from population B.
*Significant at $P < 0.01$.

detected for starch, proteins, sugars and DMC on chromosomes 10, 19, 7, and 2, respectively. These QTLs were used as examples to illustrate the kind of data obtained for the different traits. At locus 10.1.6610171, the allele C is associated with a lower starch content, whereas at locus 02.1.28100114, the allele A is associated with low dry matter content. At locus 21.1.307693 three different genotypic classes were observed and varieties homozygous for allele T (TT) had significantly higher protein content. At locus 07.1.1765193, varieties homozygous for allele T (TT) had significantly higher sugar content.

Identification of potential candidate genes

Candidate genes were detected for the majority of validated QTLs and they are presented in Table 3. The significant SNP markers were generally located in intergenic regions but four were found in intronic regions and one in an exon.

Five candidate genes were identified out of the seven QTLs for starch, including *Sucrose phosphatase 1* (EC 3.1.3.24), *Sucrose phosphatase synthase* (EC 2.4.1.14), *Isoamylase 3* (EC 3.2.1.68), *Glycosil hydrolase*, and *Serine/threonine protein kinase*.

Three candidate genes were identified out of the five QTLs for DMC, including a *Xyloglucan galactosyltransferase*, an *Ubiquitin specific protease* and an *Endo beta D Glucanase*.

A total of five candidate genes were identified out of six QTLs for sugars. They are *Pyruvate dehydrogenase kinase* (EC 1.2.4.1), *Enolase Chloroplastic* (EC 4.2.1.11), *beta glucosidase, rhamnose-galactose sugar transporter*, and *glycine cleavage system H protein*.

Finally, four candidate genes for proteins were detected out of the five QTLs (Table 3). They are a *D amino acid transaminase, Protein activity of BC1 complex kinase 7*, *Indole-3-Glycerol phosphate synthase* (EC 4.1.1.48), and *Proline rich receptor like protein kinase*.

DISCUSSION

This research provides valuable insights into the genetic architecture of key quality related traits in yam; identifies markers linked to the genomic regions controlling each trait, which is useful for MAS; validates the QTLs on a diversity panel; and identifies candidate genes from the validated QTLs.

A larger phenotypic variability was observed in population B (74F X 14 M) than in population A (74F X Kabusa) for DMC, starch, and sugars. This could be explained by greater differences observed between progenitors of this population. In both progenies the presence of transgressive hybrids (having higher or lower values than those of their parents) could be due to the heterozygosity of progenitors, and in particular of female 74F and male 14 M. Indeed, the high heterozygosity found for these two genotypes[16] would favor a high frequency of new allelic combinations, thus widening of phenotypic variation within their progenies. A similar phenomenon was observed in a previous study focusing on the same populations but on other tuber quality traits.[11]

Heritabilities of all traits were moderately high to high, with 68% to 88% of the phenotypic variation in hybrid means due to genetic differences between hybrids. This makes selection for these quality traits possible in breeding programs. The high heritability for DMC was similar to that found in 191 genotypes of different *Dioscorea* species (0.86).[12] As expected, starch content was the trait with the lowest heritabilitiy. Indeed, it is known that starch content varies between different parts of tuber ('head, middle, bottom')[18] and also during the storage of tubers.[19] Interestingly, starch was negatively correlated with proteins and sugars

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

Table 3. Quantitative trait loci validated for starch, DMC, sugar, and protein content in the diversity panel, and candidate genes identified

Trait	Chr	Pop.	LOD	R2 (%)	SNP	Alleles /Allele effect	Localization	Candidate gene
Starch	5	B	3.67	22.6	05.1_32707706**	CC/CT/TT (T-)	Intergenic	*G- type lectin Receptor like serine/threonine protein kinase*
Starch	2	B	3.63	22.4	02.1_32141722*	TG/GG (G-)		
Starch	6	B	3.39	20.1	06.1_20308422*	AA/A*/**(*+)	Intergenic	*Sucrose phosphatase 1 EC 3.1.3.24*
Starch	10	B	3.58	22.1	10.1_6610171*	T*/**(*-)	Intergenic	*Glycosil hydrolase family 5*
Starch	10	A	3.24	15.4	10.1_17600318*	CC/C* (*+)	Intergenic	*Sucrose phosphate synthase EC 2.4.1.14*
Starch	11	A	4.10	19.1	11.1_4493720**	TT/TC (C-)	Intronic	*Uncharacterized protein Loc120271825*
Starch	18	A	3.40	16.1	18.1_1695612*	GG/GA/AA (A+)	Intronic	*Isoamylase 3 chloroplastic EC 3.2.1.68*
Proteins	8	B	4.46	26.6	08.1_18931363*	CC/CG (G-)	Intergenic	*D amino acid transaminase EC2.6.1.21*
Proteins	19	B	4.33	26.1	21.1_107691*	CC/CT/TT (T+)	Intergenic	*Protein activity of BC1 complex kinase 7*
Proteins	5	B	3.85	23.6	05.1_31811203*	TT/T* (*-)	Intergenic	*Indole-3-Glycerol phosphate synthase EC 4.1.1.48*
Proteins	2	B	3.54	21.9	-	non val.		
Proteins	18	A	4.90	22.4	18.1_1629151**	GG/GA/AA (A+)	Exon	*Uncharacterized protein Loc120282302*
Proteins	10	A	3.36	16.1	10.1_16586803**	GG/GC/CC (C-)	Intergenic	*Putative proline rich receptor like protein kinase*
Sugars	9	B	4.07	24.7	09.1_19344114**	A*/** (*-)	Intergenic	*Enolase 1 chloroplastic EC 4.2.1.11*
Sugars	7	B	3.91	23.9	07.1_351152*	GG/GA/AA (A-)	Intronic	*Pyruvate dehydrogenase kinase EC 1.2.4.1*
Sugars	13	B	3.18	19.9	-	non val.		
Sugars	12	A	4.60	21.2	12.1_17583120**	TG/GG (G+)	Intergenic	*Alpha 1 Arabinofuranosidase 1 EC 3.2.1.55*
Sugars	6	A	4.49	20.7	06.1_18785797	AA/A* /** (*-)	Intergenic	*Glycine cleavage system H Protein 2*
Sugars	7	A	3.51	16.6	07.1_1765193**	CT/TT (T+)	Intergenic	*Beta glucosidase 22*
DMC	4	B	4.84	27.2	04.1_10157871*	TT/TC/CC (C-)	Intergenic	*Xyloglucan galactosyltransferase GT17*
DMC	12	B	4.77	25.1	12.1_24347534*	TT/TC/CC(C-)		
DMC	7	B	4.13	24.0	07.1_3805155*	GG/GA (A+)	Intergenic	*Endo 1,3 (4) beta D Glucanase*
DMC	1	B	3.79	22.2	01.1_601071*	GG/**/** (*-)	Intronic	*Ubiquitin like specific protease 2B*
DMC	1	A	3.00	14.3	-	non val.		
DMC	2	A	3.99	18.7	02.1_28100114**	GA/AA (A-)		

Note: *Chr* Chromosome, *Pop* Population, *SNP* marker in the QTL confidence interval that showed a significant association with phenotypic data from the diversity panel.
Note: **Significant at $P < 00.1$, * $P < 00.5$.
Note: *Non val* QTLs not validated in the diversity panel.

in both populations. This is consistent with previous results obtained on *D. alata*[9] and other root and tuber crops.[20]

The current research has led to the identification of 25 QTLs associated with the genetic variation of these four important tuber quality traits. Several QTL co-localizations were congruent with the observed genetic correlations. Three QTL co-localizations were observed for starch and proteins on chromosomes 18, 10, and 5. In addition, one QTL co-localization was observed for starch and sugars on chromosome 6. The starch QTL on chromosome 18 was located within 67 kb of the protein QTL. This short distance separating the QTLs suggests that these could be under the control of a same gene with a pleiotropic effect or several distinct, closely related genes. Starch QTLs on chromosomes 10 and 5 were at 1 Mb and 896 kb of protein QTLs, respectively. The QTL for starch on chromosome 6 was approximately 1.5 Mb from the sugar QTL. For these more distant QTLs, the hypothesis of a control by separate genes seems the most likely. Despite the existence of a negative correlation between starch and proteins, several hybrids containing both high starch and protein content were detected in both biparental populations. This can be explained by genetic recombinations between QTLs, which is feasible considering their physical distances and its telomeric chromosomal localization. Furthermore, no co-localization was observed for QTLs detected for a same trait in both populations on the same chromosome (starch on chromosome 10, sugars on chromosome 7, and DMC on chromosome 1), which is in accordance with the hypothesis that these are different loci.

Before being used for MAS, a QTL needs to be validated to confirm that its effect can be also detected in different genetic backgrounds. For this purpose, we used a contrasting diversity panel. A total of 22 QTLs could be validated showing that they are not specific to the genetic background of progenitors. These should be useful for breeding programs using MAS to select the favorable alleles and to improve yam tuber quality.

In addition, the validation process was very useful to identify the candidate genes, in particular when QTL confidence intervals were large. A total of five putative candidate genes were detected near the markers associated with starch content of which four genes (*Isoamylase ISA3, Serine/threonine protein kinase, Sucrose phosphate synthase, Sucrose phosphatase*) play important roles in the starch and sucrose metabolism. *Isoamylase ISA3* was reported to participate in the process of starch degradation in potatoes.[21] This gene was also reported to be involved in the reduction of dormancy period.[21] Both *Sucrose phosphate synthase* and *Sucrose phosphatase* play important roles in the sucrose metabolism.[22] The fourth putative gene, *Serine/threonine-protein kinase* was reported to participate in the process of starch and sugar biosynthesis in potato.[23] The fifth putative gene, *Glycosil hydrolase family*

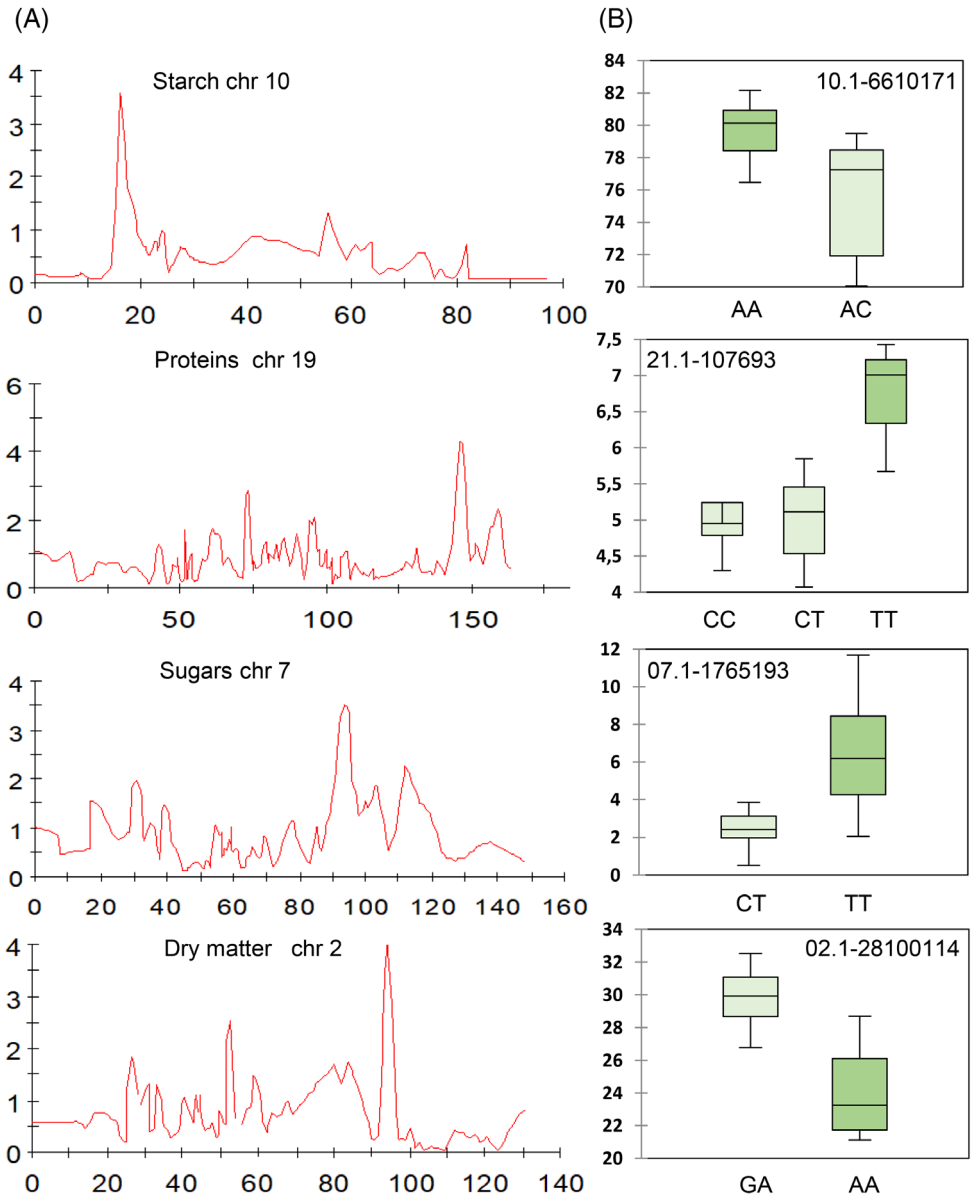

Figure 2. (A) Examples of QTLs detected in the biparental populations for starch, proteins, sugars and dry matter on chromosomes 10, 19, 7, and 2 respectively. (B) SNPs markers in the QTL confidence interval that showed a significant association with phenotypic data from diversity panel. The alleles at each locus and the allelic effects are depicted.

5, was reported to hydrolyze the glycosidic bond between two or more carbohydrates, and is present in many plant tissues.[24,25]

Five putative candidate genes were detected for sugar content, of which three play an important role in plant respiration or glycolysis (*Enolase, Pyruvate dehydrogenase kinase, Glycine H Protein*). *Pyruvate dehydrogenase kinase* was reported to be a negative regulator of the mitochondrial pyruvate dehydrogenase that plays a central role in control of cell respiration.[26] *Enolase*, also known as *Phosphopyruvate hydratase*, catalyzes the ninth and penultimate step of glycolysis. In tobacco, the glycine H protein was reported to play an important role in the photorespiratory flux.[27] Overexpression of this protein reduced the amounts of soluble sugars and increased the accumulation of starch.[27] The fourth putative gene *Beta Glucosidase* was reported to be involved in the hydrolysis of cellobiose and other oligosaccharides into glucose.[28] The fifth putative gene *UDP-rhamose/UDP-galactose transporter* was reported to be involved in the transport of nucleotide sugars (*UDP rhamose, UDP galactose*) from the cytosol into the Golgi lumen to be used in the synthesis of polysaccharides.[29]

Three putative candidate genes were identified for DMC. Of these three genes, *Xyloglucan galactosyltransferase* was reported to play a significant role in enhancing the plasticity of cell wall components through its ability to hydrolyze and reconnect the xyloglucan chains.[30] The second candidate gene, *endo Beta D Glucanase*, was reported to be involved in the cleavage of glucan chains, which are major constituents in cell walls, generating mainly oligosaccharides.[31] The third putative gene, *Ubiquitin protease*, was reported to play an important role in many plant developmental processes.[32]

Four putative candidate genes were detected for proteins. Of these four genes, *D-amino acid transaminase* was reported to play a crucial role in the biosynthesis and/or degradation metabolism

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

of different amino acids in plants, such as alanine and serine.[33] The second candidate gene, *Indole 3 glycerol phosphate synthase*, was reported to be involved in the tryptophan biosynthesis.[34] The third putative gene *Complex kinase 7* was reported to be involved in the phosphorylation of proteins both in the mitochondrial outer membrane and in chloroplasts.[35] The fourth putative gene, *Proline-rich like receptor kinase*, belongs to the hydroxyproline-rich glycoprotein (HRGP) superfamily, which was reported to be involved in many plant developmental processes.[36]

The candidate genes identified should be useful in providing a better understanding of the physiological and molecular basis of these important tuber quality traits.

AUTHOR CONTRIBUTIONS

Arnau G designed the study with support from Chair H; Ehounou AE, Maledon E, Nudol E and Leinster J contributed to sample preparation for NIRS analysis. Arnau G, Ehounou AE, Maledon E, and Nudol E contributed to sample preparation for chemical analysis. Desfontaines L and Marie-Magdeleine C generated the NIRS spectroscopic data. Arnau G performed QTL mapping, validation analysis, and candidate gene research. All authors contributed to the writing of the manuscript.

ACKNOWLEDGEMENTS

This work was supported by the RTBfoods project https://rtbfoods.cirad.fr, through a grant, OPP1178942: Breeding RTB products for end user preferences (RTBfoods) to the French Agricultural Research Centre for International Development (CIRAD), Montpellier, France, by the Bill and Melinda GatesFoundation (BMGF). A special thanks to Hernan Cerballos for his helpful comments and improvement of the manuscript and to Clair Hershey for English proofing.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1 FAO statistical database, https://www.fao.org (2022).
2 Cornet D, Sierra J, Tournebize R, Gabrielle B and Lewis FI, Bayesian network modeling of early growth stages explains yam interplant yield variability and allows for agronomic improvements in West Africa. *Eur J Agron* **75**:80–88 (2016).
3 Baah FD, Maziya-Dixon B, Asiedu R, Oduro I and Ellis WO, Nutritional and biochemical composition of D. alata (*Dioscorea* spp.) tubers. *J Food Agric Environ* **7**:373–378 (2019).
4 Egesi CN, Asiedu R, Egunjobi JK and Bokanga M, Genetic diversity of organoleptic properties in water yam (*Dioscorea alata* L). *J Sci Food Agric* **83**:858–865 (2003).
5 Otegbayo B, Tanimola A, Oroniran O, Maraval I, Forestier-Chiron N and Bugaud C, Sensory Characterization of Pounded Yam. Biophysical Characterization of Quality Traits, WP2. Iwo, Nigeria, RTBfoods Laboratory standard operating procedure, 15p (2020). https://doi.org/10.18167/agritrop/0059.
6 Adinsi L, Akissoe N, Maraval I, Forestier-Chiron N and Bugaud C, Sensory Characterization of Boiled Yam. Biophysical Characterization of Quality Traits, WP2. Cotonou, Benin,RTBfoods Laboratory standard operating procedure, 13p (2020). https://doi.org/10.18167/agritrop/00600.
7 Lebot V, Malapa R, Molisalé T and Marchand JL, Physico-chemical characterisation of yam (*Dioscorea alata* L.) tubers from Vanuatu. *Genet Resour Crop Evol* **53**:1199–1208 (2006).
8 Lebot V, Lawac F and Legendre L, The greater yam (*Dioscorea alata* L.): a review of its phytochemical content and potential for processed products and biofortification. *J Food Compos Anal* **115**:104987 (2023).
9 Lebot V and Malapa R, Application of near infrared reflectance spectroscopy for the evaluation of yam (*Dioscorea alata*) germplasm and breeding lines: quality evaluation of yam using NIRS. *J Sci Food Agric* **93**:1788–1797 (2013).
10 Ehounou AE, Cornet D, Desfontaines L, Marie-Magdeleine C, Maledon E, Nudol E *et al*., Predicting quality, texture and chemical content of yam (*Dioscorea alata* L.) tubers using near infrared spectroscopy. *J Near Infrared Spectrosc* **29**:128–139 (2021).
11 Ehounou AE, Cormier F, Maledon E, Nudol E, Vignes H, Gravillon MC *et al*., Identification and validation of QTLs for tuber quality related traits in greater yam (*Dioscorea alata* L.). *Sci Rep* **12**:8423 (2022). https://doi.org/10.1038/s41598-022-12135-2.
12 Adejumobi II, Agre PA, Onautshu DO, Adheka JG, Cipriano IM, Monzenga JC *et al*., Assessment of the yam landraces (*Dioscorea* spp.) of DR Congo for reactions to pathological diseases, yield potential, and tuber quality characteristics. *Agriculture* **12**:599 (2022). https://doi.org/10.3390/agriculture12050599.
13 Gatarira C, Agre P, Matsumoto R, Edemodu A, Adetimirin V, Bhattacharjee R *et al*., Genome-wide association analysis for tuber dry matter and oxidative browning in water yam (*Dioscorea alata* L.). *Plants (Basel)* **9**:969 (2020). https://doi.org/10.3390/plants9080969.
14 Sharif BM, Burgarella C, Cormier F, Fabien C, Mournet P, Causse S *et al*., Genome-wide genotyping elucidates the geographical diversification and dispersal of the polyploid and clonally propagated yam (*Dioscorea alata* L.). *Ann Bot* **6**:1029–1038 (2020).
15 Van Ooijen JW, *MapQTL 6, Software for the Mapping of Quantitative Trait Loci in Experimental Populations of Diploid*. Wageningen, Netherlands, Kyazma BV (2009).
16 Cormier F, Lawac F, Maledon E, Gravillon MC, Nudol E, Mournet P *et al*., A reference high-density genetic map of greater yam (Dioscorea alata L.). *Theor Appl Genet* **132**:1733–1744 (2019).
17 Tamiru M, Natsume S, Takagi H, White B, Yaegashi H, Shimizu M *et al*., Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular markerfor sex determination. *BMC Biol* **15**:86 (2017). https://doi.org/10.1186/s12915-017-0419-x.
18 Afoakwa EO, Polycarp D, Budu AS, Mensah-Brown H and Otoo E, Variability in biochemical composition and cell wall constituents among seven varieties in Ghanaian yam (Dioscorea sp.) germplasm. *Afr J Food Agric Nutr Dev* **13**:8106–8127 (2013). https://doi.org/10.18697/ajfand.59.13280.
19 Wireko-Manu FD, Oduro I, Ellis WO, Asiedu R and Maziya-Dixon B, Food quality changes in water yam (Dioscorea Alata) during growth and storage. *Asian J Agric Food Sci* **1**:66–72 (2013).
20 Lebot V, Malapa R and Jung M, Use of NIRS for the rapid prediction of total N, minerals, sugars and starch in tropical root and tuber crops. *N Z J Crop Hortic Sci* **41**:144–153 (2013). https://doi.org/10.1080/01140671.2013.798335.
21 Ferreira SJ, Senning M, Fischer-Stettler M, Streb S, Ast M, Neuhaus HE *et al*., Simultaneous silencing ofisoamylases ISA1, ISA2 and ISA3 by multi-target RNAi in potato tubers leads to decreased starch content and an early sprouting phenotype. *PLoS One* **12**:e0181444 (2017). https://doi.org/10.1371/journal.pone.0181444.
22 Fernie AR, Willmitzer L and Trethewey RN, Sucrose to starch: a transition in molecular plant physiology. *Trends Plant Sci* **7**:35–41 (2002). https://doi.org/10.1016/s1360-1385(01)02183-5 PMID: 11804825.
23 McKibbin RS, Muttucumaru N, Paul MJ, Powers SJ *et al*., Production of high-starch, low-glucose potatoes through over-expression of the metabolic regulator SnRK1. *Plant Biotechnol J* **4**:409–418 (2006).
24 Neis A and da Silva Pinto L, Glycosyl hydrolases family 5, subfamily 5: relevance and structural insights for designing improved biomass degrading cocktails. *Int J Biol Macromol* **193**:980–995 (2021). ISSN 0141-8130). https://doi.org/10.1016/j.ijbiomac.2021.10.062.
25 Minic Z, Physiological roles of plant glycoside hydrolases. *Planta* **227**:723–740 (2008).
26 Millar AH, Knorpp C, Leaver CJ, Christopher JC and Hill SA, Plant mitochondrial pyruvate dehydrogenase complex: purification and

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

identification of catalytic components in potato. *Biochem J* **334**:571–576 (1998).
27 López-Calcagno PE, Fisk S, Brown KL, Bull SE, South PF and Raines CA, Overexpressing the H-protein of the glycine cleavage system increases biomass yield in glasshouse and field-grown transgenic tobacco plants. *Plant Biotechnol J* **17**:141–151 (2019). https://doi.org/10.1111/pbi.12953 Epub 2018 Jul 22. PMID: 29851213; PMCID: PMC6330538.
28 Angkasawati VS, Kamaliyah I, Putra MD, Nata IF and Irawan C, Structural characterization and antioxidant activity of liquid sugar from Alabio potato using enzymatic hydrolysis processes. *IOP Conf Ser: Mater Sci Eng* **980**:012041 (2020). https://doi.org/10.1088/1757-899X/980/1/0120412.
29 Rautengartena C, Ebert B, Moreno I, Templec H, Hertera T, Linke B *et al.*, The Golgi UDP-Rhamnose / UDP-Galactose transporter family in Arabidopsis. *Proc Natl Acad Sci U S A* **111**:11563–11568 (2014). https://doi.org/10.1073/pnas.1406073111.
30 Baumann MJ, Eklöf JM, Michel G, Kallas AM, Teeri TT, Czjzek M *et al.*, Structural evidence for the evolution of Xyloglucanase activity from xyloglucan Endo-Transglycosylases: biological implications for cell wall metabolism. *Plant Cell* **19**:1947–1963 (2007). https://doi.org/10.1105/tpc.107.051391.
31 Perrot T, Pauly M and Ramírez V, Emerging roles of -Glucanases in plant development and adaptative responses. *Plants* **11**:1119 (2022). https://doi.org/10.3390/plants11091119.
32 Zhou H, Zhao J, Cai J and Patil SB, Ubiquitin-specific proteases function in plant development and stress responses. *Plant Mol Biol* **94**:565–576 (2017). https://doi.org/10.1007/s11103-017-0633-5.
33 Diebold R, Schuster J, Däschner JK and Binder S, The branched-chain amino acid transaminase gene family in Arabidopsis encodes plastid and mitochondrial proteins. *Plant Physiol* **129**:540–550 (2002).
34 Schlee S, Dietrich S, Kurćon T, Delaney P, Goodey NM and Sterner R, Kinetic mechanism of Indole-3-glycerol phosphate synthas. *Biochemistry* **52**:132–142 (2013).
35 Pagliarin DJ and Dixon JE, Mitochondrial modulation: reversible phosphorylation takes center stage. *Trends Biochem Sci* **31**:26–34 (2006) ISSN 0968–0004). https://doi.org/10.1016/j.tibs.2005.11.005.
36 Borassi C, Sede AR, Mecchia MA, Mangano S *et al.*, Proline-rich extensin-like receptor kinases PERK5 and PERK12 are involved in pollen tube growth. *FEBS Lett* **595**:2593–2607 (2021). https://doi.org/10.1002/1873-3468.14185.

