

# Genome-wide association studies reveal novel loci controlling tuber flesh color and oxidative browning in *Dioscorea alata*

Komivi Dossa,<sup>a,b\*</sup> Angélique Morel,<sup>a,b</sup> Mahugnon Ezékiel Houngbo,<sup>a,b,c</sup> Ana Zotta Mota,<sup>b,c</sup> Erick Malédon,<sup>a</sup> Jean-Luc Irep,<sup>d</sup> Jean-Louis Diman,<sup>d</sup> Pierre Mournet,<sup>b,c</sup> Sandrine Causse,<sup>b,c</sup> Kien Nguyen Van,<sup>e</sup> Denis Cornet<sup>b,c</sup> and Hâna Chair<sup>b,c</sup>



## Abstract

**BACKGROUND:** Consumers' preferences for food crops are guided by quality attributes. This study aimed at deciphering the genetic basis of quality traits, especially tuber flesh color (FC) and oxidative browning (OB) in *Dioscorea alata*, based on the genome-wide association studies (GWAS) approach. The *D. alata* panel was planted at two locations in Guadeloupe. At harvest, the FC was scored visually as white, cream, or purple on longitudinally sliced mature tubers. The OB was scored visually as the presence or absence of browning after 15 min of exposure of the sliced samples to ambient air.

**RESULTS:** Phenotypic characterization for FC and OB of a diverse panel of *D. alata* genotypes highlighted significant variation within the panel and across two locations. The genotypes within the panel displayed a weak structure and could be classified into three subpopulations. GWAS identified 14 and 4 significant associations for tuber FC and OB, respectively, with phenotypic variance, explained values ranging from 7.18% to 18.04%. Allele segregation analysis at the significantly associated loci highlighted the favorable alleles for the desired traits, i.e., white FC and no OB. A total of 24 putative candidate genes were identified around the significant signals. A comparative analysis with previously reported quantitative trait loci indicated that numerous genomic regions control these traits in *D. alata*.

**CONCLUSION:** Our study provides important insights into the genetic control of tuber FC and OB in *D. alata*. The major and stable loci can be further utilized to improve selection in breeding programs for developing new cultivars with enhanced tuber quality.

© 2023 The Authors. *Journal of The Science of Food and Agriculture* published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

Supporting information may be found in the online version of this article.

**Keywords:** yam; association genetics; quality traits; molecular markers; breeding

## INTRODUCTION

*Dioscorea alata* L., commonly known as greater yam or water yam, contributes significantly to food security in Africa, the Caribbean islands, and Asia.<sup>1,2</sup> Bred cultivars, with improved yields and disease resistance, have contributed substantially to meet market demands.<sup>3-5</sup> Although *D. alata* is widely adapted to different environments and shows high yield and tolerance to biotic and abiotic stresses, the quality of its tubers is lower compared to other species, such as *Dioscorea rotundata*.<sup>6</sup> Improving quality attributes of yam tubers has been largely neglected. Therefore, satisfactory end-use quality remains the primary objective of yam breeding, along with the increased yield.<sup>7</sup>

Desirable aesthetic characteristics in food depend mainly on texture and color. A recent survey by Effah-Manu *et al.*

\* Correspondence to: K Dossa, CIRAD, UMR AGAP Institut, 97170 Petit Bourg, Guadeloupe, France. E-mail: [komivi.dossa@cirad.fr](mailto:komivi.dossa@cirad.fr)

a CIRAD, UMR AGAP Institut, Petit Bourg, France

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

c CIRAD, UMR AGAP Institut, Montpellier, France

d UR1321 ASTRO Agrosystèmes tropicaux, INRAE, Petit-Bourg (Guadeloupe), Paris, France

e Plant Resources Center, Hanoi, Vietnam

emphasized tuber flesh color (FC) as a key quality trait in yams according to end-users preferences.<sup>8</sup> The tuber FC range in greater yam varies from white, yellowish, orange, pink, and purple.<sup>9</sup> However, color choice is linked to consumer preferences and the dish. For instance, the preparation of pounded yam (called 'fufu' in West Africa) requires a white to yellowish color.<sup>10</sup> In some regions, Reunion Island, Guyana, Philippine, and India, purple FC is preferred for the preparation of cakes and mouselines.<sup>11</sup> Moreover, color variation is generally attributed to differential metabolic profiles, including flavonoids, anthocyanins, and carotenoids.<sup>12,13</sup> Therefore, flesh-color variation and associated metabolites could substantially enhance the nutritional value of yam.

In yams, quality generally refers to the nutritional contents, physical appearance, and processing ease.<sup>14</sup> Oxidative browning (OB) of tubers causes undesirable color changes and can cause bitterness in pre-processed yams with an undesirable flavor.<sup>15,10,16</sup> Moreover, OB may prevent the preparation of local dishes such as pounded yam.<sup>1</sup> Although phenotypic variation concerning OB has been widely studied, the genetic factors underlying this trait have not been explored fully.<sup>17</sup> Rinaldo *et al.* showed that polyphenolic oxidation is a major cause of OB in *D. alata*.<sup>18</sup> Deciphering the genetic mechanisms underlying OB is a critical strategy to avoid discoloration in yam tubers and meet the quality requirements of consumers.

Genetic gain estimates showed that conventional breeding practices alone are insufficient to satisfy market demands.<sup>19</sup> Utilizing innovative technologies with conventional breeding can accelerate genetic gain.<sup>20</sup> Progress in genotyping techniques and the development of modern genotype–phenotype association statistical models have provided a sophisticated way forward for identifying genomic regions and molecular markers associated with phenotypic variation in crops.<sup>21</sup> These tools can be harnessed for trait introgression and rapid screening of progenies in breeding programs.<sup>21</sup> Genome-wide association studies (GWAS) have been widely adopted in plant species to uncover marker-trait associations by providing an effective way to complement the traditional quantitative trait loci (QTL) mapping.<sup>1,21</sup> Gatarira *et al.* used GWAS to identify six putative candidate genes on chromosome 5 associated with OB in *D. alata*.<sup>1</sup> Their diversity panel was limited to West Africa, hence not harboring the wealth of alleles controlling this trait. Ehounou *et al.* identified five QTLs for OB and three for tuber FC in *D. alata* using the linkage mapping approach in bi-parental populations.<sup>22</sup> However, the parents used in their study did not greatly vary for tuber FC, and their phenotyping procedures did not consider the environmental influence. It is well known that environmental conditions affect plant performance, yield, and quality, with significant implications in breeding programs.<sup>2</sup> Hence, detecting stable QTLs not influenced by environmental factors is crucial for wide-scale use in breeding programs.

The present study aimed to decipher the genetic architectures of tuber FC and OB by using a diversified panel of *D. alata* for both traits.

## MATERIALS AND METHODS

### Plant materials, growth conditions, and phenotyping

This study included 53 *D. alata* genotypes and one *Dioscorea cayenensis* clone used as an outgroup to verify the genotyping quality. This panel comprises accessions from the yam belt

countries of West Africa, the Caribbean, and the Pacific islands (Supporting Information Table S1).

The materials were planted at two locations, Roujol (16°10' 56'' N, 61° 35' 24'' W, 10 m above sea level, m.a.s.l.) and Godet (16°20' N, 61°30' 0.10 m.a.s.l.), in Guadeloupe. The average temperature and relative humidity at Godet and Roujol during the experiment were 27.1 °C and 79.5% and 25.8 °C and 87.7%, respectively. A total of 30 seedlings of each genotype were planted and split into three replicates (ten seedlings per replication). These plants were spread over three ridges (65 m long) spaced 30 cm apart within the ridges. Sugarcane straw was used as mulch over the entire plot to limit weeds at Godet, while paper mulch was used at Roujol. The plot was drip irrigated, and planting started in March 2021. Yams were planted from seeds from the harvest of year  $n - 1$  when 50% of the tubers of the genotype germinated in the storage shed. The seeds (~100 g) were treated with black soap and alcohol, by immersing them for 30 s in a mixture (10 L of water, 0.5 L of black soap, and 0.5 L of alcohol) and sown 24 h later. The tubers were harvested 7 to 9 months later, corresponding to the commercial harvest period. After harvest, tubers were stored in a storage room at 27 °C for 3 weeks.

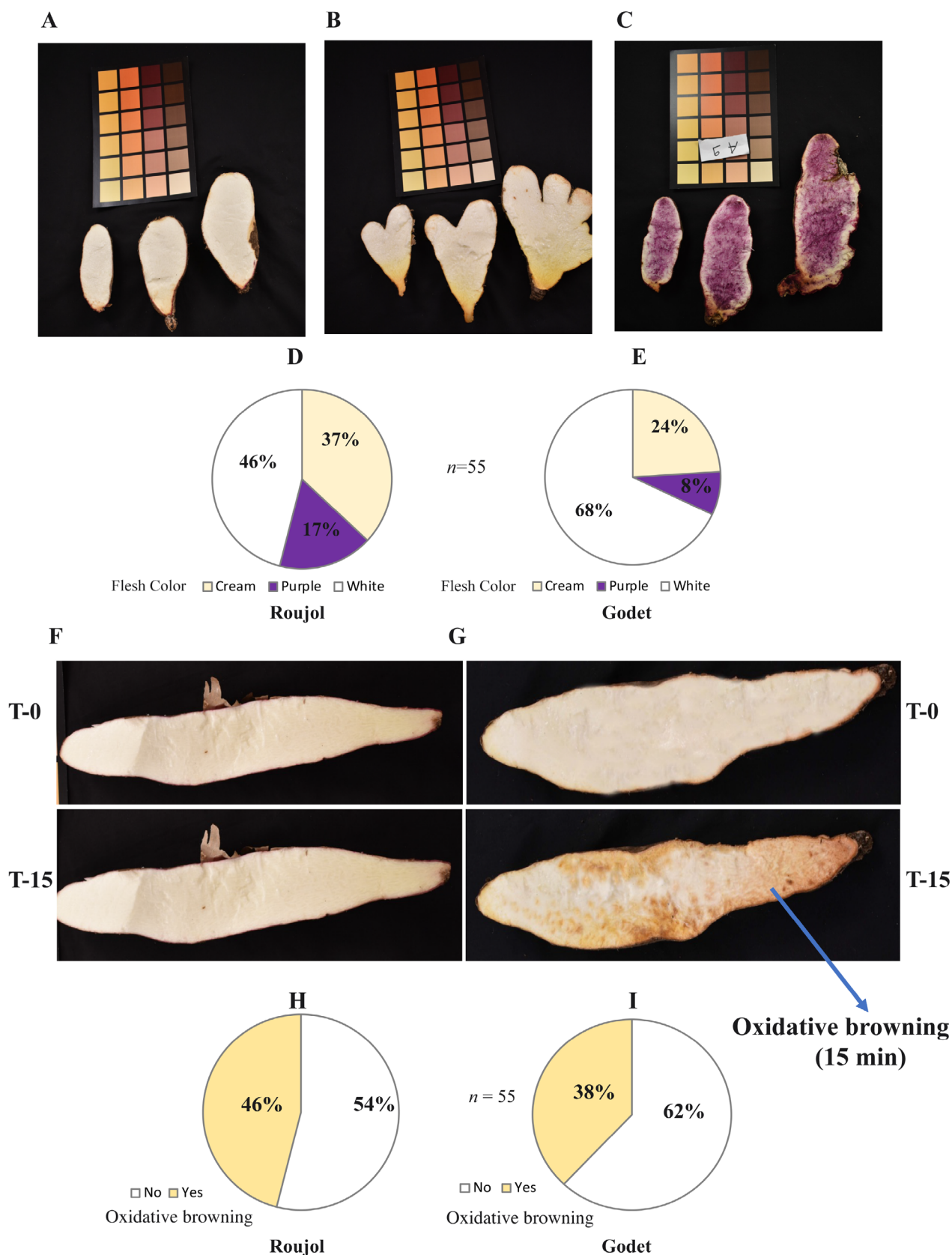
All the genotypes were characterized for phenotypic data concerning two quality traits, FC and OB. Three well-developed tubers per genotype were peeled and sliced longitudinally, and their FC was immediately scored visually as 1: white, 2: cream, or 3: purple.<sup>22</sup> The sliced samples were kept in an air-conditioned room at 25 °C and exposed to ambient air for 15 min. The OB trait was scored visually as 1: change in color/browning or 2: no change in color/no browning.<sup>22</sup> The score data were used for association analysis.

### Genotyping and single nucleotide polymorphism filtering

Genomic DNA of the 54 genotypes was extracted from the leaves using a mixed alkyltrimethylammonium bromide buffer and the kit NucleoMag Plant (Macherey-Nagel, Düren, Germany). DNA samples were quantified using ThermoFisher Scientific fluorometry Qubit (ThermoFisher Scientific, Waltham, MA, USA) and fragment length was assessed with the Agilent TapeStation system (Agilent Technologies, Santa Clara, CA, USA). Illumina Tru-SEQ DNA PCR free and TruSeq Nano DNA Preparation Kit have been applied to generate sequencing libraries. Paired high-throughput sequencing (2 × 150 bp) was performed on an Illumina NovaSeq 6000 instrument on the GeT-PlaGe platform (Toulouse, France).

The quality of the whole genome sequencing (WGS) raw reads were first checked using the FastQC package version 0.11.7,<sup>23</sup> trimmed by removing adaptor sequences and low-quality sequences using Trimmomatic version 0.39<sup>24</sup> with the following parameters: SLIDINGWINDOW, 6:15; TRAILING, 20; HEADCROP, 3; and MINLEN, 90. Clean reads were then aligned to the *D. alata* v2 reference genome using BWA-MEM version 0.7.15<sup>25</sup> with default parameters. The alignment quality was analyzed with Qualimap version 2.2<sup>26</sup> with default parameters. The calling of single nucleotide polymorphism (SNP) for each genotype was performed using GATK 4.1.6.0 with the HaplotypeCaller method.<sup>27</sup>

A consolidated variant call format (VCF) file for all genotypes was produced by VCFtools version 0.1.16,<sup>28</sup> containing about 17 million SNPs. All InDel-like variants were eliminated. Then, the VCF file was filtered using the following parameters: --max-missing 0.2 --minQ 30 --minDP 8 --maf 0.05 --allele 2. With filters applied to reduce the quantity and eliminate bad-quality SNPs, there were 1.9 million SNPs left. The SNP density was estimated using the 'CMplot' package<sup>29</sup> in R4.0.23.<sup>30</sup>



**Figure 1.** Tuber flesh color (FC) variation and oxidative browning (OB) in *Dioscorea alata* association panel. Examples of (A) white colored, (B) cream (yellow) colored, (C) purple colored tubers; variation in tuber FC within the panel at (D) Roujol, (E) Godet, where 'n' is the number of genotypes; (F) absence of OB; (G) presence of OB in a yam tuber after 15 min of cutting, where upper parts showed the color at T-0 (immediately after slicing), while lower parts show the color at T-15 (15 min after slicing); variation of OB within the panel at (H) Roujol and (I) Godet, 'No', 'Yes' represent absence and presence of OB in the genotypes, respectively.



**Population structure, principal component analysis, linkage disequilibrium, and SNP annotation**

A python centralized package named 'RESEQ2' has been designed internally in our research group, which gathers several next-generation sequencing data analysis packages (such as PLINK 1.07,<sup>31</sup> ADMIXTURE 1.23,<sup>32</sup> VCFtools 0.1.16,<sup>28</sup> GATK 4.1.6.0, ANNOVAR 2.4<sup>33</sup>). The phylogenetic relationships among the accessions were measured based on the genetic distance using the neighbor-joining method (with 1000 bootstrap replicates) in PLINK 1.07.<sup>31</sup> The resulting phylogenetic tree was visualized in MEGA-11. The population structure ( $K$  from 1–10; 100 runs per  $K$ ) was assessed with ADMIXTURE 1.23,<sup>32</sup> and the accessions with the membership coefficients ( $> 60\%$ ) were assigned to a specific Cluster. The remaining accessions were assigned to the admixture Cluster. The principal component analysis (PCA) was conducted with PLINK 1.07<sup>31</sup> with default parameters and plotted in R4.0.23. The linkage disequilibrium (LD) coefficients ( $r^2$ ) between pairwise high-quality SNPs were calculated in PLINK 1.07<sup>31</sup> with default parameters. The maximum distance between two SNPs was defined as 310 kb. The LD decay graphs were produced in R4.0.23. The annotation of the filtered VCF file was performed with ANNOVAR 2.4<sup>33</sup> with default parameters. Based on the annotation of the *D. alata* genome v2,<sup>34</sup> SNPs were categorized as occurring in coding regions (further grouped into synonymous or non-synonymous SNPs), splicing sites, untranslated regions, intronic regions, upstream and downstream regions, and intergenic regions.

**Genome-wide association studies**

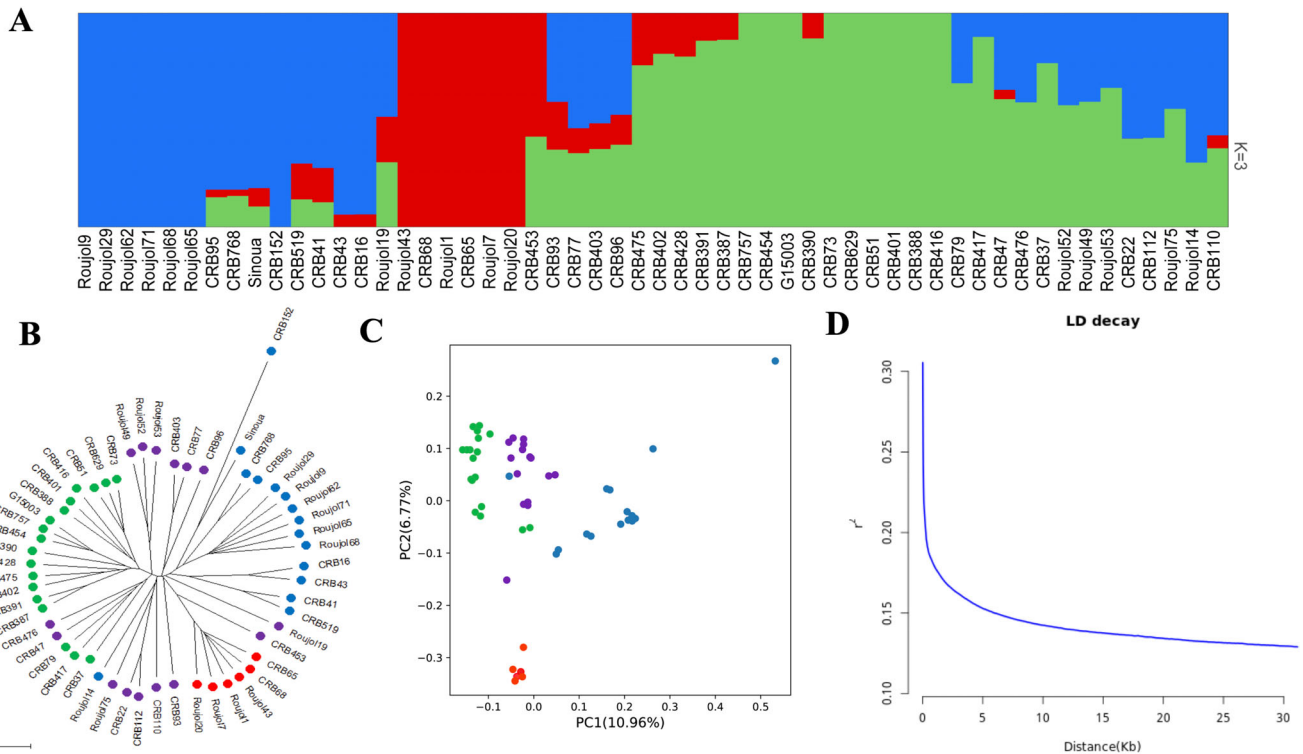
GWAS were performed with all 1.9 million SNPs using the statistical model Fixed and random model circulating probability unification (FarmCPU), and mixed linear model implemented in the GAPIT3 package<sup>35</sup> under R4.0.23.<sup>30</sup> The FarmCPU algorithm allows an improvement in the statistical power, an increase in the calculation efficiency, and a reduction of false associations in the GWAS approach.<sup>36</sup> The FarmCPU formula is:

$$Y = Si + S + e$$

$$Y = K + e$$

where  $Y$  is the phenotype,  $Si$  is the testing marker,  $S$  is the pseudo-QTN,  $K$  is the kinship matrix, and  $e$  is residual. The kinship and first principal components are fitted as covariate variables in the FarmCPU to reduce the false positives. To eliminate the ambiguity of determining associated markers that are in LD with a testing marker, the model eliminates the confounding from kinship by using only the kinship derived from the associated markers.

The Manhattan plots were also generated in R4.0.23.<sup>30</sup> SNPs having a significant association with traits were determined by the adjusted  $P$ -value. The threshold of  $P < 10^{-8}$  ( $0.05/n$ , with  $n =$  number of SNPs) was set to report a significant association. The quantile-quantile plots were generated by plotting the negative logarithms ( $-\log_{10}$ )  $P$ -values relative to their expected  $P$ -values to fit model relevance GWAS with the null hypothesis of no association



**Figure 2.** Population stratification analysis based on SNPs in *Dioscorea alata*. (A) Model-based clustering analysis of all the genotypes performed using ADMIXTURE, representing three groups and their corresponding genotypes at  $K = 3$ . (B) Circular dendrogram of the genotypes by the neighbor-joining method. *Dioscorea cayenensis* (CRB152) was used as an outgroup. Coloring followed the results from ADMIXTURE analysis, with purple-colored genotypes being the admixtures. The scale bar represents the length of the tree branch. (C) PCA based on the high-quality SNPs. Coloring followed the results from ADMIXTURE analysis with purple-colored genotypes being the admixtures. (D) LD decay estimates measured as the physical distance (kb) when LD decreased to half of its maximum value.

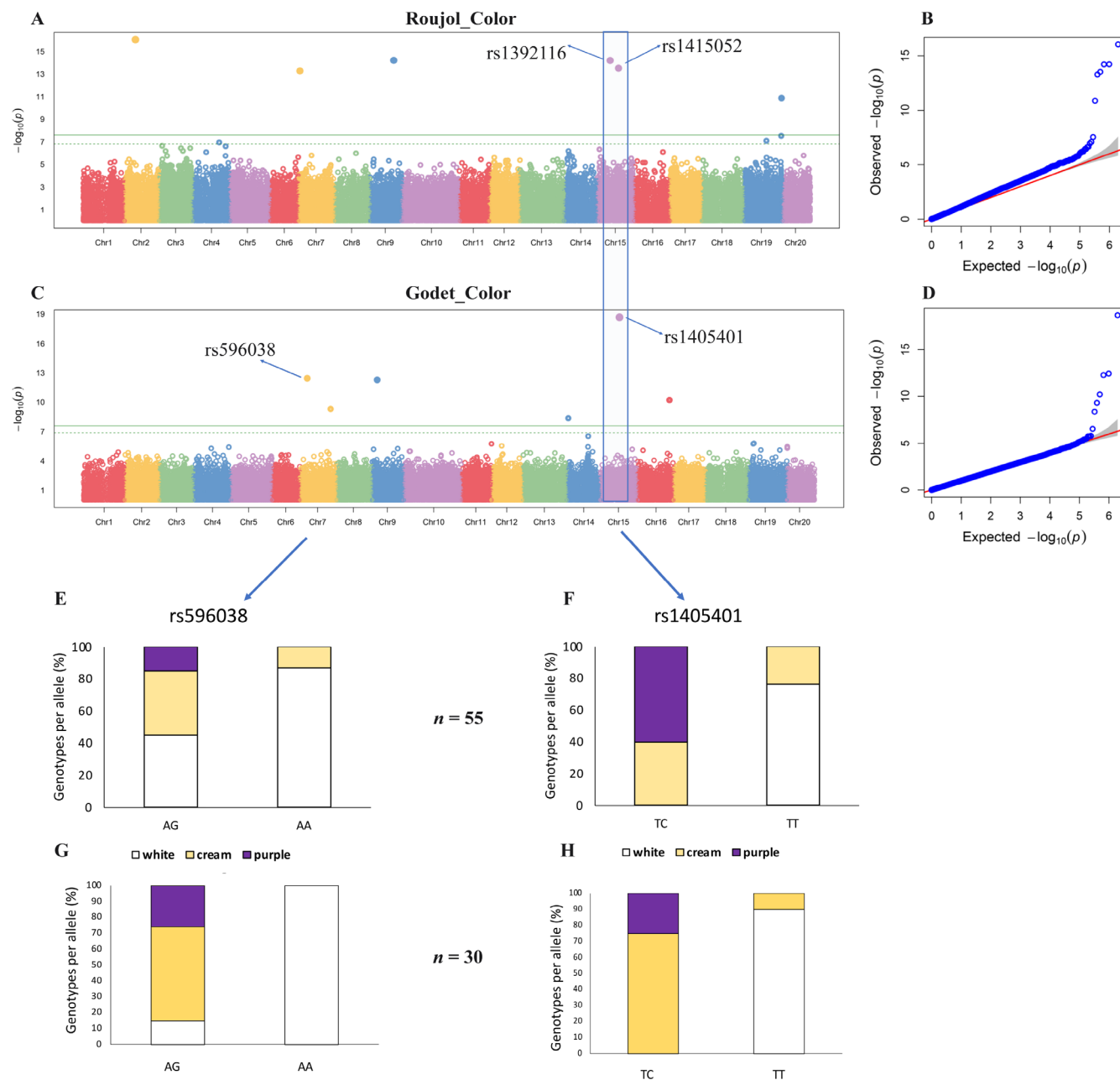
and to determine to what extent the models considered the structure of the population. The effect of alleles at significant SNPs was assessed by comparing phenotyping data for allelic groups.

### Putative candidate gene identification

To inventory potential genes near associated SNP markers for target traits, we defined a window range of the average LD distance (5 kb upstream and 5 kb downstream) and used ANNOVAR 2.4<sup>33</sup> annotated VCF file to locate genes around the significant SNPs.

### Validation of the significantly associated SNPs for tuber FC in an external panel

We obtained a phenotypic dataset on tuber FC of 30 *D. alata* accessions maintained and phenotyped in Vietnam. These accessions were planted at An Khánh District of Hoai Duc (Hanoi city) during the growing season 2016–2017. At harvest, FC of three mature tubers per accession was visually noted as explained earlier (1: white, 2: cream, or 3: purple).<sup>22</sup> The accessions were genotyped (by WGS) as explained earlier, and a VCF file was generated.



**Figure 3.** GWAS for tuber FC variation in *Dioscorea alata*. (A) Manhattan plot for tuber FC at Roujol, with the peaks indicating significant GWAS signals, and the green horizontal lines indicating the genome-wide significance threshold. (B) The QQ-plot associated with tuber FC at Roujol shows the  $-\log_{10}$  of the expected versus observed  $P$  values of each SNP (blue dots). The red line is a guide for the perfect fit to  $-\log_{10}P$ . The shaded area shows the 95% confidence interval for the QQ-plot under the null hypothesis of no association between the SNP and the trait. (C) Manhattan plot for tuber FC at Godet. (D) The QQ-plot associated with the tuber FC at Godet. (E, F) Allele segregation analysis concerning SNPs rs596038 and rs1405401, where the x-axis represents the allelic variability at each locus and y-axis represents the proportion of genotypes representing the specific allele type. Purple, cream, and white colors indicate the corresponding genotypes with purple, cream, and white FCs, respectively. (G, H) Allele segregation analysis concerning SNPs rs596038 and rs1405401 in an independent panel (size = 30 *D. alata* genotypes) grown and phenotyped in Vietnam.

Peak SNPs initially identified through GWAS were searched in the VCF file of the Vietnam panel, and alleles at these SNPs were assessed by comparing tuber FC data for allelic groups.

### In silico comparative analysis of QTLs from previous studies on tuber FC and OB in *D. alata*

QTL results from three previously published studies concerning QTLs related to tuber FC and OB in *D. alata* were extracted.<sup>1,22,34</sup> Marker sequences for each reported QTL was retrieved, and local BLAST (BLASTn) was used to identify corresponding physical positions in the reference genome v2.<sup>34</sup> The previously reported QTLs and those detected in our study were physically mapped using MapChart 2.32.<sup>37</sup>

## RESULTS

### Phenotypic characterization in the *Dioscorea alata* diversity panel

The *D. alata* panel was classified into three categories based on the tuber FC: white, cream (or yellowish), and purple FCs (Fig. 1(A)–(C)). Most accessions were white (a color generally preferred by consumers). There was variation within the two environments, suggesting that the environmental conditions influence the tuber FC in some genotypes. For instance, 46% and 68% of genotypes had white flesh tubers in Roujol and Godet, respectively (Fig. 1(D),(E)). Observed inconsistency of tuber FCs among planting sites was mainly from white to cream and *vice versa*.

Similarly, there was variation for OB in the association panel at the two locations. OB occurs after the tuber is cut and exposed to oxygen (Fig. 1(F),(G)). A total of 46% and 38% of genotypes cultivated in Roujol and Godet showed OB, respectively (Fig. 1(H),(I)). Godet might be a favorable environment for better-quality tubers.

Overall, we observed a significant variation among genotypes and environments, suggesting that the phenotypic data is suitable for GWAS.

### Genotyping data, population structure, and linkage disequilibrium

A total of 17 million SNPs were initially discovered between the genotypes of the association panel. After filtering, 1.9 million high-quality SNPs (minor allele frequency  $\geq 0.05$  and missing rate  $< 20\%$ ) were retained for genetic analysis, corresponding to a very high density of 3.9 SNPs/kb. The SNP distribution was relatively homogeneous along the chromosomes (Supporting Information Fig. S1). However, as expected, the regions near centromeres have lower densities.

Population structure analysis showed the lowest cross-validation error at  $K = 3$  (Fig. S2), indicating that the genotypes within the panel could be classified into three subpopulations, represented by the blue, red, and green colors (Fig. 2(A)). Out of the 54 genotypes, only 21 pure genotypes were identified, with the remaining classified as admixed (Table S2). Overall, the diversity panel has a weak structure, which is favorable for genetic association analysis.

Similar to the population structure results, phylogenetic and principal component analysis (PCA) did not show any obvious grouping of the studied genotypes (Fig. 2(B),(C)). The *D. cayenensis* (CRB152) used as an outgroup was clearly separated from the *D. alata* genotypes (Fig. 2(B)). PCA showed weak clustering within the population, with the principal component one (PC1) and PC2 covering only 17% of the total variation (Fig. 2(C)). Linkage disequilibrium (LD) decay provides an estimate for the resolution of association mapping to assess the optimum numbers of SNPs. LD decay was observed at 5 kb (physical distance between SNPs) (Fig. 2(D)).

Together, a large genetic diversity coupled with a low structure and a very weak LD decay were also observed in the association panel based on the SNP data, indicating that (i) the assembled panel is suitable for GWAS and (ii) a high-density SNP is necessary to obtain reliable genotype–phenotype association results.

**Table 1.** The significant association signals identified, with the allelic effects for each SNP

| No | Site   | Traits             | SNP                   | Chr   | Position   | Allele | P-Value ( $-\log_{10}$ ) | Effect (%) |
|----|--------|--------------------|-----------------------|-------|------------|--------|--------------------------|------------|
| 1  | Roujol | Flesh color        | rs125884              | Chr2  | 7 223 285  | T/C    | 16                       | 17.40      |
| 2  | Roujol |                    | rs832065              | Chr9  | 15 404 735 | G/GA   | 14                       | 13.95      |
| 3  | Roujol |                    | rs1392116             | Chr15 | 8 344 629  | C/T    | 14                       | 9.69       |
| 4  | Roujol |                    | rs1415052             | Chr15 | 13 803 586 | G/A    | 14                       | 8.41       |
| 5  | Roujol |                    | rs575623              | Chr7  | 977 139    | G/A    | 13                       | 8.87       |
| 6  | Roujol |                    | rs1884180             | Chr19 | 24 431 097 | G/A    | 11                       | 8.24       |
| 7  | Roujol |                    | rs1883574             | Chr19 | 24 226 102 | A/G    | 75                       | 7.69       |
| 8  | Godet  | Oxidative Browning | rs1405401             | Chr15 | 12 347 101 | T/TC   | 18.67                    | 18.04      |
| 9  | Godet  |                    | rs596038              | Chr7  | 4 729 592  | A/GA   | 12.45                    | 9.51       |
| 10 | Godet  |                    | rs766477              | Chr9  | 3 312 846  | T/C    | 12.29                    | 8.88       |
| 11 | Godet  |                    | rs1555677             | Chr16 | 20 775 954 | T/A    | 10.21                    | 7.42       |
| 12 | Godet  |                    | rs658251              | Chr7  | 19 942 330 | A/G    | 9.31                     | 7.18       |
| 13 | Godet  |                    | rs1275604             | Chr14 | 842 033    | A/G    | 8.37                     | 7.67       |
| 14 | Roujol |                    | rs66743               | Chr1  | 21 751 133 | C/T    | 12                       | 12.27      |
| 15 | Roujol |                    | rs1732730             | Chr18 | 20 083 173 | G/A    | 11                       | 11.53      |
| 16 | Roujol |                    | rs176324 <sup>£</sup> | Chr2  | 17 011 000 | G/T    | 85                       | 10.16      |
| 17 | Roujol |                    | rs531469              | Chr6  | 8 030 455  | G/A    | 77                       | 7.75       |
| 18 | Godet  |                    | rs176324 <sup>£</sup> | Chr2  | 17 011 000 | G/T    | 7.21                     | 10.76      |

Abbreviation: SNP, single nucleotide polymorphism;

**Table 2.** Summary of identified candidate genes associated with tuber flesh color (FC) and oxidative browning (OB) in *Dioscorea alata*

| Site   | Traits | SNP       | Position   | Upstream /downstream gene | Localization | Predicted functions  |
|--------|--------|-----------|------------|---------------------------|--------------|--|
| Godet  | Color  | rs1405401 | 12 347 101 | <i>Dioal.15G046800</i>    | intergenic   | NA   |
| Godet  |        |           |            | <i>Dioal.15G046900</i>    |              | Leucine-rich receptor-like protein kinase                                |
| Godet  | Color  | rs596038  | 4 729 592  | <i>Dioal.07G046300</i>    | intergenic   | Calcium-binding protein 39   |
| Godet  |        |           |            | <i>Dioal.07G046200</i>    |              | Auxin binding protein  |
| Godet  | Color  | rs766477  | 3 312 846  | <i>Dioal.09G024700</i>    | intergenic   | TPT Triose-phosphate Transporter family                                  |
| Godet  |        |           |            | <i>Dioal.09G024800</i>    |              | basic helix-loop-helix (bHLH) DNA-binding superfamily protein            |
| Godet  | Color  | rs1555677 | 20 775 954 | <i>Dioal.16G069700</i>    | UTR5         | P-loop containing nucleoside triphosphate hydrolases superfamily protein |
| Godet  |        |           |            | <i>Dioal.16G069700</i>    |              | P-loop containing nucleoside triphosphate hydrolases superfamily protein |
| Godet  | Color  | rs658251  | 19 942 330 | <i>Dioal.07G084000</i>    | intergenic   | Leucine-rich repeat-containing protein                                   |
| Godet  |        |           |            | <i>Dioal.07G084100</i>    |              | NB-ARC domain (NB-ARC) // Leucine-rich repeat (LRR_8)                    |
| Godet  | Color  | rs1275604 | 842 033    | <i>Dioal.14G010800</i>    | intronic     | thioredoxin-disulfide reductase  |
| Godet  | OxB    | rs66743   | 21 751 133 | <i>Dioal.01G042000</i>    | intergenic   | Phosphatidylinositol Transfer Protein Sec14p                             |
| Godet  |        |           |            | <i>Dioal.01G042100</i>    |              | Ankyrin repeat family protein  |
| Godet  | OxB    | rs1732730 | 20 083 173 | <i>Dioal.18G044800</i>    | intergenic   | Transposon protein   |
| Godet  |        |           |            | <i>Dioal.18G044900</i>    |              | Histone h1/h5  |
| Godet  | OxB    | rs176324  | 17 011 000 | <i>Dioal.02G052400</i>    | intronic     | Acireductone dioxygenase   |
| Godet  | OxB    | rs531469  | 8 030 455  | <i>Dioal.06G030700</i>    | intergenic   | NA   |
| Godet  |        |           |            | <i>Dioal.06G030800</i>    |              | Nuclear factor Y, subunit C11  |
| Roujol | Color  | rs125884  | 7 223 285  | <i>Dioal.02G036300</i>    | intergenic   | NA   |
| Roujol | Color  |           |            | <i>Dioal.02G036400</i>    |              | NA   |
| Roujol | Color  | rs832065  | 15 404 735 | <i>Dioal.09G039800</i>    | intergenic   | NA   |
| Roujol | Color  |           |            | <i>Dioal.09G039900</i>    |              | NA   |
| Roujol | Color  | rs1392116 | 8 344 629  | <i>Dioal.15G045600</i>    | intergenic   | Amino acid transporter 1   |
| Roujol | Color  |           |            | <i>Dioal.15G045700</i>    |              | NA   |
| Roujol | Color  | rs1415052 | 13 803 586 | <i>Dioal.15G047700</i>    | intergenic   | Lipid transfer protein 12  |
| Roujol | Color  |           |            | <i>Dioal.15G047800</i>    |              | NA   |
| Roujol | Color  | rs575623  | 977 139    | <i>Dioal.07G015300</i>    | exonic       | NA   |
| Roujol | Color  | rs1884180 | 24 431 097 | <i>Dioal.19G180700</i>    | exonic       | Alpha/beta-Hydrolases superfamily protein                                |
| Roujol | Color  | rs1883574 | 24 226 102 | <i>Dioal.19G177200</i>    | exonic       | Tudor/PWWP/MBT superfamily protein                                       |
| Roujol | OxB    | rs66743   | 21 751 133 | <i>Dioal.01G042000</i>    | intergenic   | NA   |
| Roujol | OxB    |           |            | <i>Dioal.01G042100</i>    |              | Ankyrin repeat family protein  |
| Roujol | OxB    | rs1732730 | 20 083 173 | <i>Dioal.18G044800</i>    | intergenic   | Transposon protein   |
| Roujol | OxB    |           |            | <i>Dioal.18G044900</i>    |              | High mobility group A  |
| Roujol | OxB    | rs176324  | 17 011 000 | <i>Dioal.02G052400</i>    | intronic     | NA   |
| Roujol | OxB    | rs531469  | 8 030 455  | <i>Dioal.06G030700</i>    | intergenic   | NA   |
| Roujol | OxB    |           |            | <i>Dioal.06G030800</i>    |              | Nuclear factor Y, subunit C11  |

Abbreviation: SNP, single nucleotide polymorphism; NA, Annotation.

### Identification of loci controlling flesh color and oxidative browning in *D. alata*

The GWAS approach was implemented with the FarmCPU model to identify the genetic variants associated with the observed variation for FC and OB, with phenotypic data from each location independently.

#### GWAS for tuber flesh color

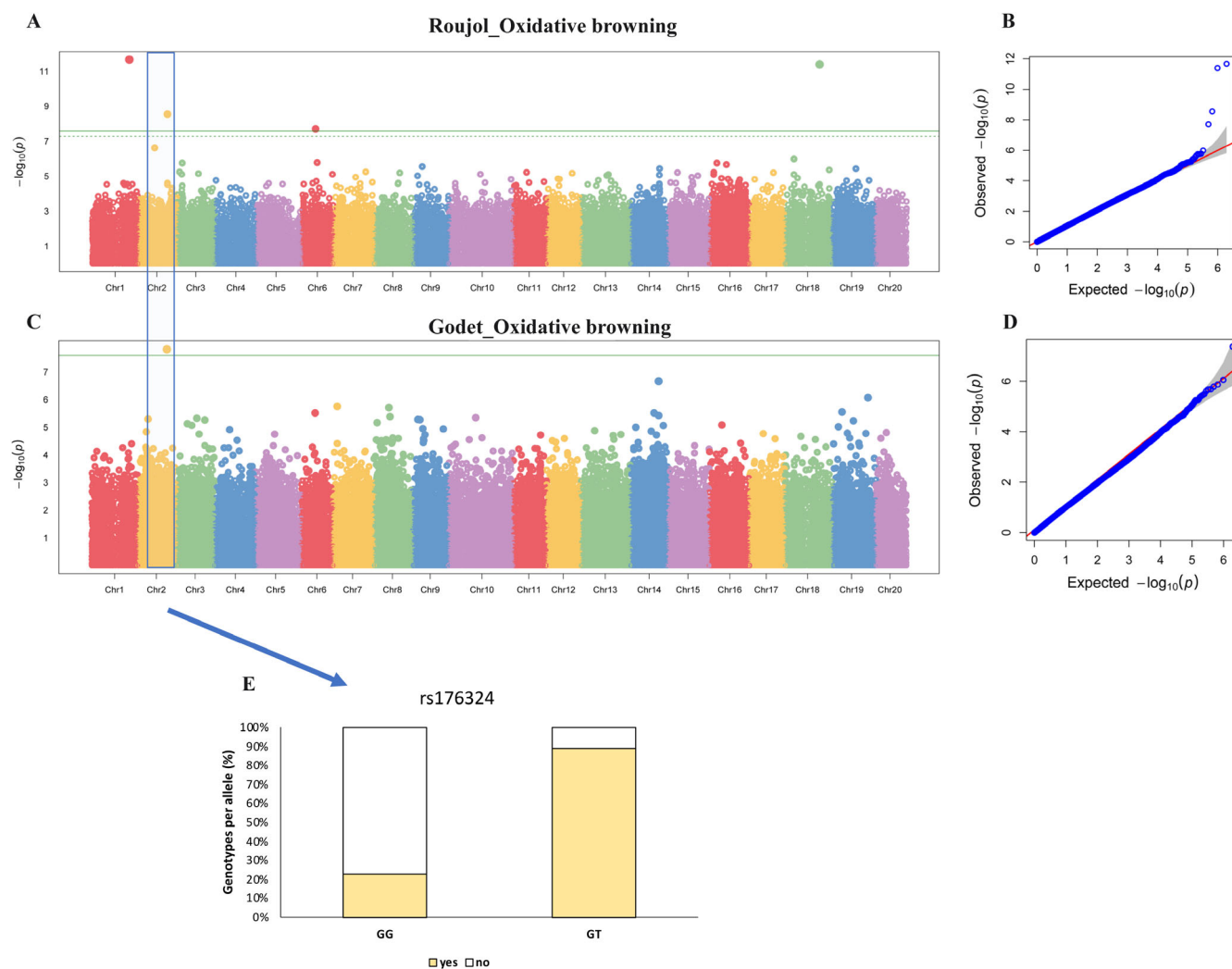
We identified 14 significant associations ( $P < 10^{-8}$ ) for tuber FC using GWAS for the two locations (Table 1). At Roujol, seven significant associations were identified on Chr2, Chr7, Chr9, Chr15, and Chr19 (Fig. 3(A),(B)), while at Godet, the significant associations were present on Chr7, Chr9, Chr14, Chr15, and Chr16 (Fig. 3(C),(D)). Only, one genomic region on Chr15 was

commonly identified at the two locations. The discrepancy in the GWAS results (for the same trait and panel) at the two locations could be attributed to the environmental effects on the phenotype. To further understand the effect of these identified loci, we determined the allelic effect for the top significant or stable SNPs (rs596038 and rs1405401) using the allele segregation analysis.

#### Allele segregation analysis for FC related peak SNPs

The two SNPs were consistently identified in close genomic regions at the two locations. The favorable alleles at each significant locus were identified, and the phenotypic variance explained was computed. For example, the detected significant SNP on Chr7 (rs596038) could explain 9.51% of the FC variation in our panel at





**Figure 4.** GWAS for OB in *Dioscorea alata*. (A) The Manhattan plot for OB at Roujol, with the peaks indicating significant GWAS signals and the green horizontal lines indicating the genome-wide significance threshold. (B) The QQ-plot associated with OB at Roujol. (C) The Manhattan plot for OB at Godet. (D) The QQ-plot plot associated with OB at Godet. (E) Allele segregation analysis concerning SNP rs176324, where cream color indicates the genotypes with OB, and the white color indicates the genotypes without OB. ‘No’, ‘Yes’ represent the absence and presence of OB in the genotypes, respectively.

Godet. We observed that the homozygous accessions (AA) at this locus have a significantly higher proportion of genotypes with white FC compared to the heterozygous accessions (AG) (Fig. 3(E)). The rs596038 is annotated as an intergenic SNP between the genes *Dioal.07G046300* (upstream) and *Dioal.07G046200* (downstream). *Dioal.07G046300* and *Dioal.07G046200* were annotated as *Calcium-binding protein 39* and *Auxin binding protein*, respectively (Table 2).

Similarly, the significant SNP on Chr15 (rs1405401) was identified with TC and TT genotypes, where 80% of the accessions with homozygous TT genotype have white FC, while heterozygous TC genotype only corresponds to the accessions with cream and purple FCs. The phenotypic variance explained by rs1405401 was estimated to be 18.04% at Godet. Moreover, we identified around the SNP rs1405401 the *Dioal.15G046900* (annotated as *Leucine-rich receptor-like protein kinase*) and *Dioal.15G046800* (no annotation) in the downstream and upstream regions, respectively (Table 2). Allele segregation analysis for the remaining SNPs also depicted allele-specific phenotypes. For instance, SNPs rs832065, rs1392116, and rs1884180 were identified with specific alleles associated with the white tuber FC (Fig. S3).

In order to validate the SNPs identified for FC and their allelic effect, we used an external panel of *D. alata* ( $n = 30$ ) planted and phenotyped in Vietnam. The independent panel globally confirmed the allelic effect of identified GWAS loci on FC, suggesting that these loci could be used to predict FC in *D. alata* (Fig. 3(G),(H)).

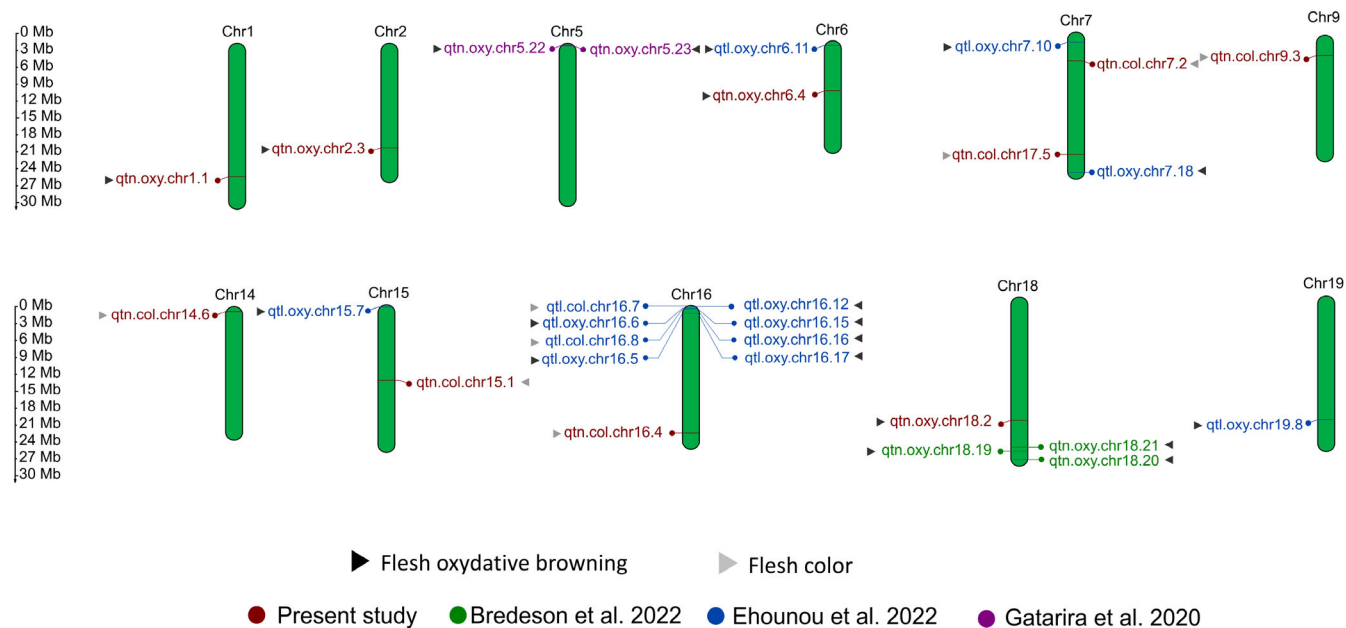
#### GWAS for oxidative browning

Four highly significant associations on Chr1, Chr2, Chr6, and Chr18 (Fig. 4(A)–(D)) for OB were found through GWAS in *D. alata*. To explore the genetic basis for tuber. Interestingly, the GWAS signal on Chr2 (rs176324) was identified at both planting locations. The SNP effect of rs176324 was estimated to be 10.16% and 10.76% at Roujol and Godet, respectively (Table 1). We further explored the alleles of rs176324 and their effect on the tuber OB.

#### Allele segregation analysis for OB-related peak SNPs

Allele segregation analysis suggested that *D. alata* genotypes with the homozygous loci GG does not display OB (Fig. 4(E)). The allele ‘T’ increases OB expression and, therefore, was considered unfavorable for breeding. The SNP rs176324 was identified





**Figure 5.** Genome-wide mapping (physical map) of quantitative trait loci from previous and current studies identified for tuber FC and OB in *Dioscorea alata*.<sup>1,22,34</sup>

in the intronic region of *Dioal.02G052400*, encoding *Acireductone dioxygenase* (Table 2). Allele segregation analysis for the remaining four SNPs also showed significant associations with OB; however, these SNPs were identified only at one planting location (Fig. S4).

We also performed GWAS using the traditional mixed linear model, however, the performance of FarmCPU was better in identifying the causal variants due to the highly heterozygous *D. alata* genome (Fig. S5). Our results confirm a previous study where authors reported that FarmCPU is more robust than the traditional mixed linear model in maize.<sup>38</sup>

### **In silico comparative analysis of genomic regions identified from previous studies on tuber flesh color and oxidative browning**

To further understand the extent of loci controlling the studied quality traits, we mapped the QTLs identified in the current study with those previously reported from three separate studies.<sup>1,22,34</sup> Mapping results showed that none of the QTLs overlapped in the four studies. Although parental genotypes used by Bredeson *et al.*,<sup>34</sup> are present in the study of Gatarira *et al.*,<sup>1</sup> and similarly, parental genotypes in the study of Ehounou *et al.*,<sup>22</sup> are present in our panel, none of the QTLs coincides, emphasizing that FC and OB, are quantitative in nature, governed by several genomic regions and affected by genotype-by-environment interactions (Fig. 5).

## **DISCUSSION**

End-user quality should be a major objective in crop breeding programs.<sup>39</sup> Unfortunately, there is a variable preference of FC in yam associated with regions and dishes, complicating breeding efforts considerably. For instance, white FC is the most preferred consumer trait in West Africa, while purple color is preferred in French Guyana, the Philippines, and India.<sup>11,40</sup> Several studies have reported significant variations in the nutritional properties of different colored yams.<sup>27,28</sup> Purple yam is

enriched in anthocyanins, which are strong antioxidants providing additional nutritional value.<sup>41</sup> As yam is widely consumed in West Africa, developing nutritionally enhanced cultivars would have a positive and significant impact on the diet of a large population. Therefore, increasing colorless anthocyanins in yam in order to maintain the desired white FC could be investigated. We observed an environmental influence on both traits, implying that the yam production environment is also important for obtaining optimal quality. Several studies have reported the environmental influence in managing yam yield,<sup>42,43</sup> however the environmental impact on yam quality traits are less documented. Identification of stable *D. alata* genotypes with desirable quality should be screened using multi-location trials for further use in breeding programs.

Recent GWAS reports in yams used panel sizes slightly higher than our study (74 and 80 accessions).<sup>44,45</sup> Yams are dieocious species with higher heterozygosity and allelic diversity as compared to hermaphrodite species. Therefore even with relatively small sample sizes, it is possible to detect SNP-trait associations, especially with qualitative or not highly quantitative traits. In the dieocious species Pistachio, only 44 accessions were used to successfully detect sex loci.<sup>46</sup> In our initial analyses, simulation on different panel sizes to estimate the power for detecting true SNP-trait associations showed that when the variance explained by the locus was superior to 7%, all panels equally performed (data not shown). Hence, with a small panel, the only risk seems to not detect true SNP-trait associations for small effect loci. This is not problematic because small effect loci are not useful in marker-assisted breeding. Moreover, with a smaller panel, statistical power could be optimized by putting more efforts into accurate phenotyping. Sharif *et al.* demonstrated that from 643 worldwide *D. alata* accessions, only 90 accessions (14%) were non-clonally related.<sup>2</sup> This means that having a large panel size dominated by true clones increases the population structure, biases estimation of allelic effects, and LD, which are crucial components for successful GWAS applications. Nonetheless,

additional efforts are needed to enlarge our diversity panel with limited clonally related accessions to improve GWAS power and detect more marker-trait associations.

Based on the high-quality resequencing data, we observed a rapid LD-decay in *D. alata*. This is expected in diecious species due to a high degree of recombination.<sup>47</sup> Our results suggest that a high SNP density is required for association analyses. Therefore, we opted for a high SNP density (1.9 million SNPs, ~3.9 SNPs/kb) in this study, which is significantly larger than any of the previous GWAS reports (less than 10 000 SNPs, ~0.02 SNP/kb).<sup>1</sup> High SNP-density with rapid LD-decay and a weak population structure have favored our GWAS.<sup>48,49</sup> Thus, 18 significant associations with moderate effects (three of which were identified overlapped in both locations and could be considered stable QTLs) have been identified. The number of QTLs identified in this study combined with previous ones<sup>1,23,34</sup> and the evidence of a strong environmental influence, indicates that FC and OB are quantitative traits. These results are similar to those from Sun *et al.*, for grape berry color.<sup>50</sup> Allele segregation analysis has been extensively used to understand the effect of significant associations on population dynamics, providing insights to select advantageous alleles.<sup>21,51</sup> In this study, allele segregation analysis at the significant loci further validated their association with the studied traits and highlighted the alleles to be considered in breeding programs.

A total of 24 genes around the significant GWAS signals, 14 for tuber FC and ten for OB have been identified. *Dioal.07G046300*, *Dioal.07G046200*, and *Dioal.15G046900* encoding *Calcium-binding protein 39*, *Auxin binding protein*, and *Leucine-rich receptor-like protein kinase*, respectively, were identified as putative candidate genes for tuber FC. Pandey *et al.*, demonstrated that *Calcium-binding protein* is involved in pigmentation in tobacco.<sup>52</sup> Similarly, *Auxin binding protein* plays a significant role in plant tissue coloration.<sup>53</sup> Interestingly, Hazak *et al.*, discussed that *Calcium-binding protein* and its signaling pattern coincides with the expression pattern of auxin-regulated genes.<sup>54</sup> We hypothesize that *Dioal.07G046300* and *Dioal.07G046200*, located on the upstream and downstream regions of rs596038 on chromosome 15, might act harmoniously to regulate FC in *D. alata*. However, the underlying mechanisms are still not clear. Further functional study can elaborate on their potential roles in obtaining the desirable FC.

OB is a major quality issue in yam. Previous reports suggested that polyphenol oxidation (PPO) contributes to OB in yam.<sup>16,18</sup> A recent study by Gatarira *et al.* identified several candidate genes associated with OB on chromosome 5.<sup>1</sup> Contrary to their report, we did not find any significant association for OB on chromosome 5; instead, a stable signal on chromosome 2 was identified. Another study by Ehounou *et al.* used QTL analysis and identified several QTLs related to OB on chromosomes 5, 7, 15, 16, and 19.<sup>22</sup> Bredeson *et al.* identified three peroxidase-encoding genes (*Dioal.18G098800*, *Dioal.18G099400*, and *Dioal.18G100900*) on chromosome 18 using bi-parental mapping populations.<sup>34</sup> The obvious differences between previous studies and our results are probably due to the different phenotyping procedures, genotyping methods and the quantitative nature of the trait. For example, two main phenotyping approaches (chromameter measurements<sup>18</sup> and a visual scoring<sup>22</sup>) for yam tuber FC and OB have been used in the past. However, the correlation between these methods has not been investigated properly. We identified *Dioal.02G052400* encoding *A. dioxygenase* as a putative candidate gene for tuber OB. *Acireductone dioxygenase* is a metal-binding metalloenzyme that binds with Fe<sup>2+</sup> or Ni<sup>2+</sup> in the methionine salvage pathway.<sup>55</sup> Another report emphasized the role of

*A. dioxygenase* in ethylene biosynthesis,<sup>56</sup> which is a key regulator for OB in crops through the modulation of PPO and other antioxidant enzymes.<sup>57</sup> How ethylene potentially modulates polyphenol profiles leading to tuber OB in *D. alata*<sup>18</sup> warrants further study.

Taken together, our study provided novel insights into the genetic control of FC and OB in *D. alata* tubers. Accessions accumulating favorable alleles at the different significant loci are valuable sources for future breeding programs. Nonetheless, our study showed that marker assisted breeding for these two quality traits could be difficult. This is because numerous loci with moderate contributions are involved and there is a strong environmental influence. The potential of genomic selection on predicting FC and OB in *D. alata* should be investigated. Furthermore, we identified novel putative candidate genes for both traits. Functional analysis of these genes will contribute to a better understanding of their implications in FC and OB in *D. alata*.

## ACKNOWLEDGEMENTS

The authors are grateful to Marie-Claire Gravillon, Christophe Perrot and Elie Nudol for their contribution to the phenotyping. The editorial comments by Hernán Ceballos improved greatly the quality of this manuscript. This work was supported by the CGIAR Research Program on Roots, Tubers and Bananas (CRP-RTB) and the grant opportunity INV-008567 (formerly OPP1178942): Breeding RTB Products for End User Preferences (RTBfoods), to the French Agricultural Research Centre for International Development (CIRAD), Montpellier, France, by the Bill & Melinda Gates Foundation (BMGF): <https://rtbfoods.cirad.fr>.

## AUTHOR CONTRIBUTIONS

Conceptualization: Komivi Dossa, Hàna Chair. Data curation and formal analysis: Komivi Dossa, Angélique Morel, Mahugnon Ezékiel Hougbo, Ana Zotta Mota, Erick Malédon, Jean-Luc Irep, Jean-Louis Diman, Sandrine Causse, Pierre Mournet, Kien Nguyen Van, Denis Cornet. Funding acquisition: Denis Cornet, Hàna Chair. Writing – original draft: Komivi Dossa. Writing – review and editing: Denis Cornet, Hàna Chair. All authors have read and approved the final version of this manuscript.

## EDITORIAL POLICIES AND ETHICAL CONSIDERATIONS

All the experimental methods including field studies were performed in accordance with relevant guidelines and regulations.

## CONFLICT OF INTEREST STATEMENT

The authors declare no competing financial interests.

## DATA AVAILABILITY STATEMENT

The Illumina NovaSeq 6000 sequencing raw data are available in the NCBI Sequence Read Archive, under the BioProject number: PRJNA880983. Data will be released upon publication of this manuscript. The phenotypic datasets are available from the corresponding author upon request.

## SUPPORTING INFORMATION

Supporting information may be found in the online version of this article.

## REFERENCES

- 1 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.). *Plan Theory* **9**:969 (2020).
- 2 Sharif BM, Burgarella C, Cormier F, Mournet P, Causse S, Van KN *et al.*, Genome-wide genotyping elucidates the geographical diversification and dispersal of the polyploid and clonally propagated yam (*Dioscorea alata*). *Ann Bot* **126**:1029–1038 (2020).
- 3 Srivastava AK, Gaiser T, Paeth H and Ewert F, The impact of climate change on yam (*Dioscorea alata*) yield in the savanna zone of West Africa. *Agr Ecosyst Environ* **153**:57–64 (2012).
- 4 Pouya N, Hgaza VK, Kiba DI, Bomisso L, Aighewi B, Aké S *et al.*, Water yam (*Dioscorea alata* L.) growth and tuber yield as affected by rotation and fertilization regimes across an environmental gradient in West Africa. *Agronomy* **12**:792 (2022).
- 5 Cornet D, Marcos J, Tournèze R and Sierra J, Observed and modeled response of water yam (*Dioscorea alata* L.) to nitrogen supply: consequences for nitrogen fertilizer management in the humid tropics. *Eur J Agron* **138**:126536 (2022).
- 6 Otoo E, Opoku-Agyeman M, Dansi A, Aboagye L, Acheremu K and Tetteh J, Increasing farmers and breeders access to yam (*Dioscorea* Spp) diversity: the case of forest-savannah transition agroecology. *Afr J Agric Res* **10**:772–782 (2015).
- 7 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).
- 8 Effah-Manu L, Wireko-Manu FD, Agbenorhevi JK, Maziya-Dixon B, Oduro IN and Baah-Ennum TY, The effect of gender on end-user preferences for yam quality descriptors. *Food Sci Nutr* **10**:3890–3904 (2022).
- 9 Clawson DL, Harvest security and intraspecific diversity in traditional tropical agriculture. *Econ Bot* **39**:56–67 (1985).
- 10 Champagne A, Legendre L and Lebot V, Chemotype profiling to guide breeders and explore traditional selection of tropical root crops in Vanuatu, South Pacific. *J Agric Food Chem* **57**:10363–10370 (2009).
- 11 Lebot V, Tropical root and tuber crops. Cabi (2019).
- 12 Wang A, Li R, Ren L, Gao X, Zhang Y, Ma Z *et al.*, A comparative metabolomics study of flavonoids in sweet potato with different flesh colors (*Ipomoea batatas* (L.) Lam). *Food Chem* **260**:124–134 (2018).
- 13 Tanaka Y, Sasaki N and Ohmiya A, Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J* **54**:733–749 (2008).
- 14 Lebot V, Malapa R, Molisalé T and Marchand J-L, Physico-chemical characterisation of yam (*Dioscorea alata* L.) tubers from Vanuatu. *Genet Resour Crop Evol* **53**:1199–1208 (2006).
- 15 Shittu TA and Olaitan OF, Functional effects of dried okra powder on reconstituted dried yam flake and sensory properties of ojojo—a fried yam (*Dioscorea alata* L.) snack. *J Food Sci Technol* **51**:359–364 (2014).
- 16 Martin FW and Ruberte R, Polyphenol of *Dioscorea alata* (yam) tubers associated with oxidative browning. *J Agric Food Chem* **24**:67–70 (1976).
- 17 Srivichai S and Hongsprabhas P, Profiling anthocyanins in Thai purple yams (*Dioscorea alata* L.). *Int J Food Sci* **2020**:1–10 (2020).
- 18 Rinaldo D, Sotin H, Pédro D, Le-Bail G and Guyot S, Browning susceptibility of new hybrids of yam (*Dioscorea alata*) as related to their total phenolic content and their phenolic profile determined using LC-UV-MS. *LWT* **162**:113410 (2022).
- 19 Ray DK, Ramankutty N, Mueller ND, West PC and Foley JA, Recent patterns of crop yield growth and stagnation. *Nat Commun* **3**:1–7 (2012).
- 20 Voss-Fels KP, Herzog E, Dreisigacker S, Sukumaran S, Watson A, Frisch M *et al.*, “SpeedGS” to accelerate genetic gain in spring wheat. *Appl Genet Genomic Res Cereals*:303–327 (2019).
- 21 Nazir MF, He S, Ahmed H, Sarfraz Z, Jia Y, Li H *et al.*, Genomic insight into the divergence and adaptive potential of a forgotten landrace *G. áhirsutum* L. *purpurascens*. *J Genet Genomics* **48**:473–484 (2021).
- 22 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**:1–14 (2022).
- 23 Andrews S, FastQC: a quality control tool for high throughput sequence data, in *Babraham Bioinformatics*. Babraham Institute, Cambridge, United Kingdom (2010).
- 24 Bolger AM, Lohse M and Usadel B, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120 (2014).
- 25 Li H, Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. *Bioinformatics* **28**:1838–1844 (2012).
- 26 Okonechnikov K, Conesa A and García-Alcalde F, Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**:292–294 (2016).
- 27 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A *et al.*, The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**:1297–1303 (2010).
- 28 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158 (2011).
- 29 Yin L, CMplot: circle manhattan plot. R package version 3 (2020).
- 30 Team RC, R Foundation for Statistical Computing; Vienna: 2018. *R: a language and environment for statistical computing*[Google Scholar] (2016).
- 31 Bender D, Maller J, Sklar P, de Bakker P, Daly M and Sham P, PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* **81**:559–575 (2007).
- 32 Alexander DH, Novembre J and Lange K, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**:1655–1664 (2009).
- 33 Wang K, Li M and Hakonarson H, ANNOVAR: functional annotation of genetic variants from next-generation sequencing data. *Nucleic Acids Res* **38**:1–7 (2010).
- 34 Bredeson JV, Lyons JB, Oniyinde IO, Okereke NR, Kolade O, Nnabue I *et al.*, Chromosome evolution and the genetic basis of agronomically important traits in greater yam. *Nat Commun* **13**:1–16 (2022).
- 35 Wang J and Zhang Z, GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genomics Proteomics Bioinf* **19**:629–640 (2021).
- 36 Liu X, Huang M, Fan B, Buckler ES and Zhang Z, Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet* **12**:e1005767 (2016).
- 37 Voorrips R, MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* **93**:77–78 (2002).
- 38 Cui Z, Dong H, Zhang A, Ruan Y, Jiang S, He Y *et al.*, Denser markers and advanced statistical method identified more genetic loci associated with husk traits in maize. *Sci Rep* **10**:1–10 (2020).
- 39 Randhawa HS, Bona L and Graf R, Triticale breeding—progress and prospect, in *Triticale*. Springer, Switzerland, pp. 15–32 (2015).
- 40 Otegbayo B, Madu T, Oroniran O, Chijioke U, Fawehinmi O, Okoye B *et al.*, End-user preferences for pounded yam and implications for food product profile development. *Int J Food Sci Technol* **56**:1458–1472 (2021).
- 41 Chen X, Sun J, Zhu Q, Xiao Y, Zhang H, Huang Y *et al.*, Characterizing diversity based on phenotypes and molecular marker analyses of purple yam (*Dioscorea alata* L.) germplasm in southern China. *Genet Resour Crop Evol* **69**:1–13 (2022).
- 42 Asfaw A, Standard operating protocol for yam variety performance evaluation trial, Ed (2016).
- 43 Norman PE, Tongoona PB, Danquah A, Danquah EY, Agre PA, Agbona A *et al.*, Genetic analysis of agronomic and quality traits from multi-location white yam trials using mixed model with genomic relationship matrix. *Global J Bot Sci* **10**:8–22 (2022).
- 44 Mondo JM, Agre PA, Asiedu R, Akoroda MO and Asfaw A, Genome-wide association studies for sex determination and cross-compatibility in water yam (*Dioscorea alata* L.). *Plan Theory* **10**:1412 (2021).
- 45 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).
- 46 Kafkas S, Ma X, Zhang X, Topçu H, Navajas-Pérez R, Wai CM *et al.*, Pistachio genomes provide insights into nut tree domestication and ZW sex chromosome evolution. *Plant Commun* **4**:100497 (2022).
- 47 Barnaud A, Laucou V, This P, Lacombe T and Doligez A, Linkage disequilibrium in wild French grapevine, *Vitis vinifera* L. subsp. *silvestris*. *Heredity* **104**:431–437 (2010).
- 48 Migicovsky Z, Gardner KM, Money D, Sawler J, Bloom JS, Moffett P *et al.*, Genome to phenome mapping in apple using historical data. *Plant Genome* **9**:1–15 (2016).
- 49 Sul JH, Martin LS and Eskin E, Population structure in genetic studies: confounding factors and mixed models. *PLoS Genet* **14**:e1007309 (2018).
- 50 Sun L, Li S, Jiang J, Tang X, Fan X, Zhang Y *et al.*, New quantitative trait locus (QTLs) and candidate genes associated with the grape berry

- color trait identified based on a high-density genetic map. *BMC Plant Biol* **20**:1–13 (2020).
- 51 Liu Y, Nazir MF, He S, Li H, Pan Z, Sun G *et al.*, Deltapine 15 contributes to the genomic architecture of modern upland cotton cultivars. *Theor Appl Genet* **135**:1401–1411 (2022).
- 52 Pandey GK, Pandey A, Reddy VS, Deswal R, Bhattacharya A, Upadhyaya KC *et al.*, Antisense expression of a gene encoding a calcium-binding protein in transgenic tobacco leads to altered morphology and enhanced chlorophyll. *J Biosci* **32**:251–260 (2007).
- 53 Tan L, Tan C, Zuraida A, Hossein H, Goh H, Ismanizan I *et al.*, Overexpression of auxin binding protein 57 from Rice (*Oryza sativa* L.) increased drought and salt tolerance in transgenic *Arabidopsis thaliana*, in *IOP Conference Series: Earth and Environmental Science*. IOP Publishing, UK 012038 (2018).
- 54 Hazak O, Mamon E, Lavy M, Sternberg H, Behera S, Schmitz-Thom I *et al.*, A novel Ca<sup>2+</sup>-binding protein that can rapidly transduce auxin responses during root growth. *PLoS Biol* **17**:e3000085 (2019).
- 55 Liang S, Xiong W, Yin C, Xie X, Jin Y-J, Zhang S *et al.*, Overexpression of OsARD1 improves submergence, drought, and salt tolerances of seedling through the enhancement of ethylene synthesis in rice. *Front Plant Sci* **10**:1088 (2019).
- 56 Guo T, Zhang X, Li Y, Liu C, Wang N, Jiang Q *et al.*, Overexpression of MdARD4 accelerates fruit ripening and increases cold hardiness in tomato. *Int J Mol Sci* **21**:6182 (2020).
- 57 Jung S-K and Watkins CB, Involvement of ethylene in browning development of controlled atmosphere-stored 'Empire' apple fruit. *Postharvest Biol Technol* **59**:219–226 (2011).