Genomic insights into greater yam tuber quality traits

Ana Zotta Mota[1,2,3†], Komivi Dossa[1,4†], Mathieu Lechaudel[5,6], Denis Cornet[1,2], Pierre Mournet[1,2], David Lopez[1,2], Hana Chaïr[1,2]

[1]CIRAD, UMR AGAP, F34398-Montpellier, France,
[2] AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro,Montpellier, France,
[3] Institut Sophia Agrobiotech, Sophia Antipolis, France
[4]CIRAD, UMR AGAP Institut, 97170 Petit Bourg, Guadeloupe, France.
[5]UMR Qualisud, CIRAD, F97130-Capesterre-Belle-Eau, Guadeloupe, France
[6]QualiSud, Université Montpellier, Institut Agro, CIRAD, Avignon Université, Université de la Réunion, 34398 Montpellier, France

[†]These authors contributed equally to this work
Corresponding author: hana.chair@cirad.fr

ABSTRACT

Yams are important tubers widely consumed in developing countries in various forms, mainly boiled, pounded and fried. Tuber quality is a key criterion for acceptance by the various actors in the value chain. However, the genetics of yam tuber quality has not yet been comprehensively investigated. Given this lack of knowledge, we used population genomics and candidate gene association approaches to unravel the genetic basis of the three main quality attributes of boiled yam, namely texture, starch content and colour indices. We re-sequenced the whole genome of 127 yam genotypes with different quality attributes, and performed an enrichment of the already available functional genome annotation using comparative genomics. Population genomics analysis highlighted three main genetic groups and 1,053 genes under selection. We focused this work on three known quality trait-related pathways: pectin, starch content and flavonoid biosynthesis, and inventoried in the genome all the structural genes related to those pathways by comparative genomics. We improved the functional annotation of the three pathways with 48% more genes. A total of 18 candidate genes containing genetic variants significantly associated with the target traits were detected, including eight genes that were also found under selection. The texture-related candidate genes were distributed between the pectin and starch pathways. Overall,

the use of comparative genomics has allowed the generation of an unprecedented genomic resource. The improved functional annotation of the yam genome is a promising predictive tool for identifying other core genes associated with any trait of interest to meet the growing need and diversified demands for yams.

Keywords: comparative genomics, texture, pectin, starch, flavonoids, *Dioscorea alata*

INTRODUCTION

Releasing varieties with enhanced quality is a target in the breeding process, and failure to achieve it can lead to the rejection of improved varieties. However, food quality is a complex attribute that reflects the preferences of all actors in the value chain: producers, processors, retailers and consumers. In recent years, there has been an increasing amount of studies on the identification of genomic regions associated with quality traits (Liao et al., 2021; Xia et al., 2020), with a focus on economically important crops such as rice, tomato, citrus, etc. (Xiao et al., 2021; Dominguez et al., 2020; Butelli et al., 2019). Underutilised crops such as tropical root and tuber crops have been largely neglected. However, root and tuber crops are among the most important crops for subsistence and commercial purposes after cereals.

Yams, of the Dioscoreaceae family, are important edible tuber crops, mainly in developing countries. The most widespread species is greater yam (*Dioscorea alata* L.), cultivated mainly for its starchy tubers in tropical regions (Wu et al., 2014; Siqueira et al., 2014; Girma et al., 2014 Lebot et al., 1998). Greater yam displays a wide tolerance to different environments (Orkwor and Asadu, 1998). It is a dioecious and autopolyploid species (2n = 2x = 40, 3x = 60 and 4x = 80) (Arnau et al., 2009). The recent publication of the genetic map of this species (Cormier et al., 2021) followed by the release of its reference genome (Bredeson et al., 2022) have paved the way to work on the genetic architectures of the traits of interest. Genomic regions associated to anthracnose resistance, tuber oxidative browning, tuber morphological attributes, and dry matter have been revealed (Bredeson et al., 2022; Gatarira et al., 2020; Ehounou et al., 2022). Although it is well known that tuber quality is important for varieties adoption in root and tuber crops

65 (Dufour et al. 2021), the genetics of yam tuber quality has not yet been comprehensively
66 investigated.

67 Greater yam is consumed in various forms, mainly boiled, pounded, fried or baked (Honfozo et
68 al., 2021). The quality of boiled yam, as characterised by sensory testing with consumers, is related
69 to white colour, crumbly, sticky to the fingers, non-fibrous and easy to chew texture, sweet taste
70 and good smell (Honfozo et al., 2021). The choice of colour and the consistency of the tubers after
71 cooking are essential characteristics that have been previously selected in other crops, such as
72 cassava and potato (Alves-Pereira et al., 2022; Hardigan et al., 2017). Texture, which determines
73 the ability of the raw material to soften after cooking while maintaining its firmness, is a complex
74 trait. The main factors influencing texture are starch content and its distribution in the tuber, cell
75 wall structure and composition, and degradation of the middle lamella of the cell wall (reviewed
76 in Taylor et al., 2008; Ross et al., 2011). The cell wall is composed of 90-95% polysaccharides,
77 divided into cellulose, hemicellulose and pectin, and only 5-10% protein. Cell wall
78 polysaccharides and starch content are mainly associated with two metabolic pathways: pentose
79 and glucuronate interconversions, and starch and sucrose metabolism. Colour, one of the most
80 important varietal rejection factors, is partly determined by genes in the flavonoid biosynthetic
81 pathway (He et al., 2021). Despite their importance for the main quality attributes of boiled yam,
82 these pathways have been very little studied.

83 Over the last decade, hundreds of genomes have been released. Likewise, much efforts have been
84 put on improving the genome annotation in order to extract biological knowledge from genomic
85 sequences (Marks et al., 2021). However, annotating a genome is time and resources consuming.
86 Automatic pipelines can produce inaccurate genome annotation and their results often require
87 manual curation (Bolger et al., 2017). In non-model species, this can be especially true, since
88 several genes lack a precise functional annotation, relying on public databases and gene similarity
89 only. The use of comparative analysis allows to find homologous and orthologous genes from
90 different species and within the species of interest (Mota et al., 2020), and predict more precisely
91 their functions. Using the annotation produced with non-manual curated methods (similarity-
92 based), together with comparative genomics methods can increase the number of genes with a
93 functional annotation, and also help to better assign the functional annotation of other genes.

94 In this study, we re-sequenced the whole genome of 127 yam genotypes with different quality
95 traits from distinct geographical locations. We focused our work on the three metabolic pathways
96 involved in texture and in starch content (pentose and glucuronate interconversions, starch and
97 sucrose metabolism), and in colour (flavonoid biosynthesis). Through the use of comparative
98 genomics, we were able to improve the annotation of a number of genes, including those in the
99 three major metabolic pathways targeted, thereby enriching the annotation of the *D. alata* genome.
100 We conducted a population genomics study to identify genes under selection. To complement our
101 results, we performed a candidate gene association analysis using phenotypic data on texture,
102 colour and starch content. We identified candidate genes associated with key quality traits for
103 boiled yam and assessed their expression profiles using transcriptomic analysis. Overall, we
104 generated a consistent and valuable amount of genomic resources for future research on important
105 agronomic traits in yam.
106

107 # RESULTS

108

109 **Pattern of genome-wide variation, population structure and signatures of selection**

110

111 We generated the whole-genome sequences of 127 genotypes of *Dioscorea* spp. from diverse
112 origins (Africa, Caribbean, Pacific and Asia) (Figure 1A) using short-read sequencing with a mean
113 coverage of 37X (Supplemental Figure 1, Supplemental Table 1). The polymorphism analyses in
114 *D. alata* genotypes yielded a total of 63 M SNPs after filtering. The type of SNPs was divided as
115 2.8% of variants present in exons, 1.3% of synonymous variants, and 48% intergenic variants
116 (Figure 1B, Supplemental Figure 2). For diversity purposes, only diploid genotypes (107) were
117 used (Supplemental Table 1), since the diversity of the diploid genotypes represents the overall
118 diversity of the species as demonstrated previously (Sharif et al., 2020). Three major groups were
119 separated phylogenetic tree (Supplemental Figure 3A) and by principal component analysis
120 (PCA). A first group consisting mainly of Asian genotypes, a second group composed of Pacific
121 genotypes, and a final group of African and Caribbean genotypes (Figure 1C), which agrees with
122 our previous study, performed by the genotype-by-sequencing analysis on a larger *D. alata*
123 population (Sharif et al., 2020). The genetic structure separated the genotypes into four clusters

124 (Figure 1D). However, as the fourth group had only six genotypes, a robust statistical analysis
125 could not be conducted. Consequently, we conducted our further analysis with *K=3*. The cluster
126 A composed mainly of Pacific representatives and, to a lesser extent, African and Caribbean
127 genotypes, while the cluster B was composed mostly by representatives from Africa and
128 Caribbean, and finally the cluster C had mostly representatives from Asia (26 out of 30) (Figure
129 1D, Supplemental Figure 3B), which is in accordance with the PCA results. The nucleotide
130 diversity (π) was low in all gene pools $0.407e^{-4}$. The π values were significantly different between
131 the clusters, with the highest values obtained for the cluster A and C ($\pi = 0.399e^{-4}$ and $0.396e^{-4}$,
132 respectively) and the lowest for the Cluster B ($\pi = 0.33e^{-4}$).
133 To evaluate the genetic variance among the three clusters, we calculated the fixation index values
134 (*Fst*) for each region in a window of 50kb and a step-window size of 10kb. We found in total 2,141
135 outlier regions *i.e.,* with significant values of *Fst* (varying from $7.85e^{-06}$ to 0.64) (Figure 2A, B and
136 C). The comparison of Cluster B *vs* C, showed the highest values of *Fst* and the highest number
137 of outlier regions (881), followed by the comparison of A *vs* C (789) then A *vs* B (471) (Figure
138 2D). These outlier regions harboured a total of 1,053 unique genes (Figure 2D). We found 11
139 outlier regions in common among the three genetic clusters, which could be the result of an early
140 purifying selection in *D. alata*. Only four genes were present in these 11 common outlier regions:
141 two closely located genes on chromosome 1 and two closely located genes on chromosome 4. The
142 genes found on chromosome 1 were both annotated as non-specific serine/threonine kinase
143 (EC:2.7.11.1). Similarly, both genes on chromosome 4 are annotated as proteasome endopeptidase
144 complexes (EC:3.4.25.1). Functional analysis of the 1,053 genes based on Gene Ontology (GO)
145 terms showed that various functional pathways were enriched. The most enriched GO terms were
146 tripeptide transmembrane transporter, transmembrane receptor protein binding, stamen filament
147 development, hormone-mediated signalling pathway and lignin biosynthetic process (Figure 2E).
148
149 **Search of genes involved in pathways associated with tuber quality using combined**
150 **strategies**
151
152 To better understand the genetic basis of *D. alata* tuber quality, we focused on the genes described
153 to be involved in one of the three pathways targeted (pentose and glucuronate interconversions,
154 starch and sucrose metabolism, and flavonoid biosynthesis). To obtain the most complete set of

155 genes from these pathways, we used two complementary strategies: a keyword search and a
156 comparative genomics approach. For the first strategy, genes annotated on the *D. alata* reference
157 genome (Bredeson et al., 2022) with the enzyme commission numbers (EC number) in one of the
158 three pathways were selected. In total, we could retrieve 322 genes from these pathways
159 encompassing 177 EC numbers (Figure 3A; Supplemental Table 2). The pentose and glucuronate
160 interconversions pathway include most genes (150) and EC numbers (73), followed by starch and
161 sucrose metabolism (136 genes, 76 EC numbers) and flavonoid biosynthesis metabolism (83
162 genes, 28 EC numbers). To complement the set of annotated genes, a comparative genomics
163 approach was employed using a comparison of 45 diverse plant genomes (Mota et al., 2020)
164 (Supplemental Table 3). Groups of orthologous predicted proteins from different plants (OG),
165 were used to infer the function of proteins that were not annotated by other methods. Since the
166 clustering of proteins in OG does not depend on their previous function annotation but rather on
167 their protein sequence, this method can uncover missing genes from a keyword search. The
168 orthology analysis resulted in a total of 58,916 OG from 1,706,645 proteins (weblink dataverse
169 Cirad). Out of the total proteins analysed, 93.7% could be assigned to an OG. We identified 3,982
170 OG that are commonly found in all 45 species, despite their significant evolutionary distance. As
171 expected, species that are closely related either phylogenetically or by genus have a greater number
172 of OG in common than distantly ones (Supplemental Figure 4A, B). When we narrowed our
173 comparison to monocotyledons (which comprised 18 species), we found a total of 5,544 OG that
174 are commonly shared among the species. Only three OG were not present in *D. alata* (Figure 3B).
175 The low number of missing groups in *D. alata* species was also observed by Bredeson et al. (2022),
176 when comparing it with species exclusively from the Dioscoreales family.
177 Among the five Dioscoreales species, we found 8,029 OG in common. *Trichopus zeylanicus*
178 showed the highest number of exclusive OG (1,951) compared to the other *Dioscorea* species (*D.*
179 *dumetorum*, *D. rotundata*, *D. zingiberensis*, and *D. alata*) and *D. alata* the lowest one (336) (Figure
180 3C). Additionally, *D. alata* and *D. rotundata* had a high similarity, and likely underwent the same
181 duplication events (Bredeson et al., 2022). We identified 14 OG that comprise 163 exclusive
182 proteins of Dioscoreales (Supplemental Table 4), with most of these proteins (39) belonging to *D.*
183 *alata*. Based on their EC numbers, most of these OG are hydrolases. They are not exclusive to
184 *Dioscorea* species but their sequence conservation among other species is low, resulting in
185 exclusive OG (*i.e.,* species-specific).

186 Once we identified all the OG that included representatives from *D. alata*, we narrowed these
187 groups down to the three pathways targeted in this study. For most of the OG found in the three
188 pathways, at least one gene from each of the 45 species was present, or the majority of them
189 (weblink dataverse Cirad; Supplemental Figure 4B). These OG were not exclusive to any
190 particular phylogenetic clade, genus, or species, indicating the importance of these proteins in
191 biological processes for all plant species. Although present in all species, the number of proteins
192 for each OG was uneven, which could be the result of both bias originating from assembly and
193 annotation discrepancies, as well as duplication events that occurred after the differentiation of
194 certain species (in-paralogs) (Pont et al., 2011).
195 The comparative analysis was not used to increase the number of EC numbers for these pathways
196 but to complete the set of previously annotated genes (Bredeson et al., 2022). We could retrieve
197 new protein coding genes from the comparative genomics analysis which were not found from the
198 keyword search. Thus, we found additional 137 genes for the pentose and glucuronate
199 interconversions pathway, 12 genes for starch and sucrose metabolism pathway and 46 genes for
200 flavonoid biosynthesis metabolism (Figure 3A; Supplemental Figure 5, 6, 7). We were able to
201 associate one orthologous group to each EC number, which contains one or more genes, except
202 for the pentose and glucuronate interconversions pathway, for which the 73 EC numbers were
203 collapsed in only 57 orthogroups. For the other two pathways, each EC number corresponds to an
204 orthologous group. After combining these two approaches, we could retrieve a total of 564 genes
205 from 161 orthogroups for the three pathways, representing 48% more genes than using only one
206 strategy. These genes were distributed along all the chromosomes of *D. alata* with chromosomes
207 2, 5, 11, and 17 having the highest number of genes and chromosome 16 the lowest (Figure 1B).
208 To determine the polymorphism within the genes found in the three pathways, we selected the
209 open read frame regions since we are only interested in the genic polymorphisms in this study. A
210 total of 406,325 SNPs on 564 candidate genes, from the three pathways was selected. By keeping
211 only SNPs annotated as high impact (*i.e.* non-synonymous, stop codon gain, frame shift)
212 responsible for the most significant differences on the protein sequences, a final number of 4,287
213 SNPs was used for further analysis, with the pathway of pentose and glucuronate interconversions
214 having the highest number (2,912), followed by starch and sucrose pathway (765) and flavonoid
215 biosynthesis (610).
216

217 **Identification of genes under selection in the three targeted pathways**

218

219 Once we have determined the complete set of *D. alata* genes belonging to the three pathways, we
220 searched for regions under selective pressure, using *Fst* analysis. When comparing the 564 genes
221 from the three pathways to the 1,053 genes from the *Fs*t analysis, we found 27 genes in common
222 including nine genes found exclusively through the comparative genomic approach (Figure 3D,
223 Supplemental table 5). The majority of the 158 SNPs on these 27 candidate genes were on the
224 pentose and glucuronate interconversions pathway (119 polymorphisms in 10 genes) only found
225 between Cluster B (mostly African genotypes) and C (Asian genotypes), followed by starch and
226 sucrose (23 polymorphisms in 7 genes) and flavonoids (16 polymorphisms in 10 genes).
227 On the pentose and glucuronate interconversions pathway, we found three pectate lyase genes
228 (*Dioal.01G025700.1*, *Dioal.01G025800.1*, *Dioal.01G035100.1*), which together with the
229 polygalactunorase gene (PG) (*Dioal.06G005800.1*) and pectin methylesterase gene (PME)
230 (*Dioal.08G105700.1*) are known as pectin-degrading enzymes involved in the
231 demethylesterification of homogalacturonans (Leng et al., 2017). PME family acts prior to
232 hydrolysis by PG enzyme that degrades the homogalacturonate backbone (polygalacturonate).
233 Pectate lyase gene family is responsible for the cleavage of pectate, which is the product of pectin
234 degradation. From the same pathway, we also found four malectin-like receptor-like kinase genes
235 (MLD) under selection (*Dioal.20G055000.1*, *Dioal.20G055100.1*, *Dioal.20G055200.1*,
236 *Dioal.20G055300.1*). The MLD genes, frequently erroneously described as Receptor-Like
237 Kinases, can bind to pectin, with a preference for more highly methyl esterified pectin, causing a
238 loosening in the cell-wall (Li et al., 2016).
239 On the starch and sucrose pathway, seven genes were under selection. Among them, two cellulose
240 synthase genes (*Dioal.08G106000* and *Dioal.08G130500*), were under selection between clusters
241 B and C. Additionally, we identified a starch synthase gene (*Dioal.09G032200*) and a nudix-
242 hydrolase (*Dioal.11G000200*), under selection among A and C and among A and B, respectively,
243 a 1,3-beta-glucan synthase (*Dioal.10G014800*), among A and B and a trehalose-6-phosphate
244 phosphatase (*Dioal.04G184200*) under selective pressure among A and C genetic clusters. Those
245 genes are responsible for starch synthesis in plants (Kolbe et al., 2005; Muñoz et al., 2006; Keeling
246 and Myers, 2010). Trehalose phosphate phosphatase is involved in the synthesis of Trehalose
247 consisting of two 1,1-linked glucose molecules (Kolbe et al., 2005). Lastly, we found a beta-

248 amylase gene (*Dioal.04G168200*) under selection. The beta-amylase gene family was previously
249 described to increase the firmness of cooked sweet potatoes when its enzyme activity is low (Banda
250 et al., 2021).
251 On the flavonoids pathway were found under selection 2-oxoglutarate-dependent dioxygenase
252 (*Dioal.01G016100* and *Dioal.01G016200*) and chalcone isomerase (*Dioal.08G131200.1*). The
253 involvement of these genes was mostly associated with plant-pathogen interaction, and their
254 relationship with quality traits remains to be investigated (Fang et al., 2022).
255
256 **Candidate gene association studies for quality traits in *D. alata***
257
258 To get a deeper insight in the involvement of the genes annotated from the three metabolic
259 pathways on the tuber quality, we conducted a candidate gene association study. We used
260 phenotypic data collected from two-years field trials on 45 genotypes among the 125 *D. alata*
261 genotypes. They were selected to cover the genetic diversity and to minimize the genetic structure.
262 A set of 10 traits related to starch content, colour indices (Brown Index, the Hue for purple and the
263 Browning over time) and texture were measured. Texture was measured with two different types
264 of probes. A conical probe was used to measure the total area (N sec), which reflects the total force
265 required to penetrate the sample at a constant speed. A flat plate probe was used during a double
266 compression cycle to measure hardness, gumminess, springiness, and cohesiveness. Descriptive
267 statistics of the various tuber quality traits evaluated in this study are shown in Supplemental Table
268 6. The colour indices BI, HI exhibited wide range of variation (12.57-143 and -11.99-99.92,
269 respectively), showing the presence of varying coloured yam tubers in the association panel. The
270 tuber SC was on average 77.93±2.10 with the minimum (69.46 %) and maximum (81.26 %)
271 observed in CRB417 and Roujol9, respectively. The textural traits also varied extensively within
272 the panel, indicating varying boiled yam qualities of the accessions (Supplemental Table 6).
273 Taking together, our results indicate a wide variation in the tuber quality traits within the 45
274 genotypes and pinpoint that these traits are largely quantitative in nature.
275 Based on the general linear model, we obtained significant associations between SNPs and the
276 phenotypic data. We detected a total of 22 quantitative trait nucleotides (QTNs) distributed on 10
277 chromosomes, for seven out of the ten traits evaluated in this study (Table 1). Hereafter, we present
278 in more details five key genes associated with textural traits, starch content and colour. The

279 expression profiles of the candidate genes were assessed in six phenotypically diverse genotypes
280 using transcriptomic analysis.
281 Out of the 22 QTNs, 17 were detected for texture traits with fourteen and three within pentose and
282 glucuronate interconversions, and sucrose and starch pathways, respectively (Table 1 and
283 Supplemental Figure 8). Within pentose and glucuronate interconversions pathway, we found
284 pectin esterase (*Dioal.19G187400.1*) and xylose isomerase genes (*Dioal.13G053400*) associated
285 to springiness and gumminess. Leucine-rich repeat (*Dioal.10G024400)*, protein kinases
286 (*Dioal.02G062700*, *Dioal.02G062800*), a malectin-like encoding gene (*Dioal.20G055400*) were
287 associated to springiness and had a stop gain caused by a transition and by a transversion. Two
288 other malectin-like genes found under selection were associated to hardness. Within starch and
289 sucrose pathway we found three candidate genes associated to gumminess, total area and hardness,
290 α-amylase (*Dioal.07G051400.1*), 4-α-glucanotransferase *(Dioal.07G067100)* and beta-
291 glucosidase (*Dioal.13G093500*), respectively.
292 Pectin methylesterase (*Dioal.04G130600.1*), was associated to gumminess and total area and
293 found under selection. A tuber pectin methyl esterase activity has been identified as a potential
294 factor impacting cooked potato textural properties (Ross et al. 2011). The SNP S4_20715506 (C/T)
295 located in the only exon of *Dioal.04G130600.1* (Figure 4A), results in an amino acid change (H to
296 W), with a potential impact on the enzyme activity. We observed that the C-allele at S4_20715506
297 decreases gumminess and total area (Figure 4B and C). None of the accessions used for
298 transcriptome analysis had the T-allele, hence we were unable to evaluate the expression pattern
299 of this gene between two haplotypes. Nonetheless, our results suggest that allelic variation in
300 *Dioal.04G130600.1* probably affects the enzyme activity, leading to varying textural properties of
301 boiled yam.
302 Beta-glucosidase, associated to hardness, is a cellulase enzyme playing a part in metabolism of
303 cell wall polysaccharides (Sampedro et al., 2017). The alternative C-allele at the qH13.2 had a
304 strong effect on hardness (62%) by modifying the splice donor on position 837 in the gene
305 *Dioal.13G093500* (Table 1). The genotypes CRB47, Roujol49, and Roujol75 which are
306 homozygous for the G-allele showed an overall higher expression than the genotypes with the C-
307 allele (CRB96, Roujol62, and Roujol9).
308 The variation of the starch content in the association panel was associated with three QTNs (Table
309 1). They were annotated as 1,3-beta-glucan-synthase (*Dioal.10G014800*), beta-amylase

310 (*Dioal.15G119500*) and glucose-1-phosphate adenyl transferase (*Dioal.17G112400*). The 1,3-
311 beta-glucan-synthase had the highest effect on the starch content and was also found in an outlier
312 region by our *Fst* analysis (Supplemental Table 5). A recent study demonstrated that manipulation
313 of this gene induces different starch and polysaccharide profiles in barley grain (Garcia-Gimenez
314 et al. 2020). The genotypes harbouring the G-allele had a premature stop codon instead of a
315 glycine. They had 14% less starch content compared to the genotypes with the T-allele. We
316 speculated that the gain of a stop codon results in a non-functional or altered protein which affects
317 the SC.

318 Lastly, within the flavonoid biosynthesis pathway, we identified two QTNs qHI4.1 and qBI19.1
319 for Brown index (BI) and HUE index (HI), respectively (Table 1). The gene *Dioal.19G083800.1*
320 encodes for flavanone 3-hydroxylase (F3H). It catalyses the 3-beta-hydroxylation of 2S-
321 flavanones to 2R,3R-dihydroflavonols which are intermediates in the biosynthesis of major
322 pigments (flavonols, anthocyanidins, and proanthocyanidins) (Falcone Ferreyra et al., 2012).
323 *Dioal.19G083800.1* contains SNP S19_17029453 (G/C) in the last exon, with the G-allele leading
324 to a premature stop codon, a truncated protein and likely a non-functional enzyme (Figure 4A).
325 S19_17029453 was significantly associated with BI. The genotypes with the G-allele tend to have
326 brown colour in contrast to the genotypes with the C-allele displaying mostly whitish tuber flesh
327 (Figure 4D). Thus, this mutation may block the formation of flavonoid compounds and redirect
328 the phenylpropanoid pathway to the formation of other brownish compounds like oxidized
329 phenolic acids. Expression pattern of *Dioal.19G083800* between two contrasting accessions with
330 G/C alleles showed that it was expressed nearly 20 times higher in heterozygous (Roujol75; C/G)
331 than in homozygous (Roujol62; G/G) genotypes (Figure 4E). Overall, these results suggest that a
332 functional mutation in *Dioal.19G083800* partly modulates colour formation in yam tuber.

333 Another key gene detected for colour variation (purple and non-purple yam varieties) was
334 *Dioal.04G073400*, encoding for UDP-glucuronosyltransferase (UGT). Anthocyanins are unstable
335 water-soluble pigments. UGTs are key enzymes that stabilizes anthocyanidin by attaching sugar
336 moieties to the anthocyanin aglycone (Li et al. 2016). In this study, the T-allele at the SNP
337 S4_16031453 leads to a stop codon in the middle of *Dioal.04G073400* (Figure 4A), which will
338 probably result in a non-functional enzyme. Intriguingly, all accessions with the T-allele displayed
339 purple tuber (very low HI score) (Figure 4F). This could be explained by the fact that non-

340 functionality of an enzyme producing uncoloured flavanones could redirect the pathway to
341 anthocyanins biosynthesis resulting in the purple colour of the tubers.

342 Altogether, the candidate gene association analysis showed that variants in several genes affect the
343 functions and potentially the expression of key structural genes involved in the regulation of yam
344 tuber quality traits. Interestingly, only 38% of the genes found in this analysis were retrieved from
345 the keyword annotation, while the remaining were retrieved by comparative genomics.

346

347 **DISCUSSION**

348 In this study, the use of comparative genomics and genome sequence analysis of 127 *Dioscorea*
349 *alata* accessions enabled the identification of candidate genes in the main metabolic pathways
350 involved in yam tuber quality. The functional annotation of the *D. alata* genome remains
351 suboptimal since there are few genomic data available for this crop (Chaïr et al., 2022). Only
352 recently a complete *D. alata* genome reference was published (Bredeson et al., 2022). Therefore,
353 the comparative genomics combined with keyword search used in this study, allowed the
354 identification of novel genes associated with the three metabolic pathways targeted, pentose and
355 glucuronate interconversions, starch and sucrose metabolism, and flavonoid biosynthesis
356 pathways. We found more paralogs for each metabolic pathway, even if the number of ECs
357 remained the same. We annotated orthologous groups using the conserved domains, which
358 facilitated the correction of incomplete or outdated gene assignments, such as for Malectin-like
359 genes. In addition, based on the orthology analysis we found gene structure errors, such as for
360 OG0082990 and OG0083049 (weblink CIRAD dataverse), which calls for manual curation of
361 important gene families to further improve the functional annotation of *D. alata* genome.

362 We found that various gene ontology categories were enriched from the genes under selection
363 between the three identified genetic groups. The most enriched pathways under selection were
364 transmembrane transport, protein phosphorylation, response to wounding, hormone signalling,
365 lignin biosynthetic process, known to play important physiological roles in defence response
366 during plant-stress interactions (Zhang and Sun, 2021). This indicates most probably a selection
367 for adaptation to biotic and abiotic conditions. The downsizing to the three pathways revealed

368 under selection cellulose synthase, required for primary and secondary cell wall cellulose synthesis
369 (Richmond and Somerville, 2000), pectate lyase and PG known as pectin degrading enzymes.
370 Additionally, we also identified genes involved in starch synthesis, nudix hydrolase, soluble starch
371 synthase 3 and 1-3 beta-glucan synthase (Miao et al., 2017), confirming the importance of starch
372 content in yam during farmer selection and most probably during domestication (Meyer and
373 Purugganan, 2013). Genes for the pentose and glucuronate interconversion pathway were found
374 only between African and Asian genetic groups. This result deserves further investigation to
375 understand this exclusivity. Five genes from the pectin and starch-sugar pathways were found in
376 common between the *Fst* analysis and candidate gene association study, showing that the quality
377 of yam tubers was also a factor of differentiation and selection during yam evolution.

378 The candidate gene association analysis enabled the identification of fourteen pectin and three
379 starch candidate genes associated with texture, confirming the involvement of the cell wall and
380 starch content in boiled yam texture (Kouadio et al., 2013). Among the seventeen candidate genes,
381 three have high effect and are involved directly in the cell wall composition. Beta-glucosidase
382 explains more than 60% of the hardness confirming it as a major QTN in boiled yam texture. Its
383 role in cell wall thickness has been demonstrated in different crops such as potato (Ross et al.,
384 2011) or barley in which it participates in endosperm cell wall degradation during germination
385 (Leah et al., 1995). In cotton (*Gossypium hirsutum*), the overexpression of the *GhBG1A* a gene
386 encoding beta-glucosidase repressed fiber length but promoted cellulose biosynthesis resulting in
387 thicker fiber cell wall (Watcharatpong et al., 2020). The four remaining predicted genes with high
388 effect on texture parameters were involved in pectin pathway. Pectin methylesterase, found also
389 under selection, encodes for an enzyme which plays an important role in both pectin remodelling
390 and disassembly and consequently in firming and softening of cell wall (Atmodjo et al., 2013). In
391 our study, it has a negative effect on both total area and gumminess indicating its involvement in
392 pectin degradation during yam boiling, such as in cooked potato (Ross et al., 2011). In cassava, its
393 involvement in the root softening process during cassava retting was demonstrated in addition to
394 pectin/pectate lyase and polygalacturonase genes (Ngolong Ngea et al., 2016). These two last
395 genes were found under selection but not associated with any trait investigated. Finally, Xylose
396 isomerase seems to play an important role in boiled yam texture. It was associated with gumminess

397 and springiness most probably by catalysing the reversible isomerisation of pentoses such as
398 arabinose, one of the components of pectin side chain (Mu et al., 2018).

399 Such as for cereals and other tuber crops, the total starch content of yams is a primary determinant
400 of tuber quality. Preferred mealy cooking yam cultivars had significantly higher starch content
401 (Kouadio et al., 2013). Among the three genes associated with starch content, the 1,3-beta-glucan
402 synthase had the highest effect. It is often considered to be a cellulase family member, plays an
403 important role in cellulose structure. In this study this gene family was found under selection and
404 predicted to be associated with starch content. The (1,3;1,4)-beta-D-glucans are most abundant in
405 walls of the cereals, specifically in the starchy endosperm of grain, where they can contribute up
406 to 70% by weight of the walls in barley, rye, and oats (Burton and Fincher, 2012). Whether it is
407 related to starch content on yam or other tuberous starchy crops remains unclear.

408 The colour is a determinant key trait in yam varieties adoption. While in West Africa, white colour
409 is preferred, in the Pacific or Caribbean colourful plates are more appreciated. In our study the
410 flavanone 3-hydroxylase (F3H) was found involved in brown index with a high effect. This gene
411 is involved in the accumulation of catechins in tea plants (*Camelia sinensis* L.) (Singh et al., 2008).
412 In carnation flower (*Dianthus caryophyllus* L.) it has been found involved in colour and fragrance
413 (Zuker et al., 2002). RNAi-Mediated silencing of the F3H confirmed that this gene is one of the
414 key enzymes required for the biosynthesis of flavonoids in strawberry fruit (Jiang et al., 2013).
415 The UDP-glucosyltransferase was found associated with purple colour confirming its role in
416 colour formation in yam tuber. None of these genes were found by Wu et al. (2015) through
417 transcriptome analysis of two contrasting *D. alata* genotypes for tuber purple colour, or by a
418 genome-wide association analysis (Dossa et al., 2023). These results indicate that more genes
419 could be involved in this trait. Moreover, we mainly focused our study on SNPs located in coding
420 regions, and only structural genes involved in three targeted pathways were investigated. This
421 approach therefore excludes the regulatory genes such as transcription factors as well as other
422 candidate genes and genome components that control agronomic traits. Similarly, we did not
423 observe significant gene expression changes between contrasting genotypes for the all identified
424 candidate genes, probably because coding SNPs mainly affect protein functions but at a less extent
425 transcript levels.

426 In conclusion, quality is one of the main criteria selected during evolution and adaptation, which
427 is reflected in the genes identified in this study. We found twenty-two candidate genes, in pentose
428 and glucuronate, sucrose and starch and flavonoid biosynthesis pathways, associated to the three
429 main attributes of boiled yam: texture, starch content and tuber colour. This is the first work that
430 considers the different cell wall constituents and their effect on texture. In addition, we were able
431 to confirm the expression of five candidate genes in these pathways. Further validation of these
432 results on a larger panel of associations by performing a genome-wide association analysis
433 combined with metabolomic and transcriptomic analyses will confirm the robustness of this
434 approach. Validation of the identified genes will pave the way for favourable allele pyramiding in
435 breeding programmes. Moreover, the use of comparative genomics and the keyword search
436 approach to retrieve genes from the three pathways has proved effective. We have been able to
437 retrieve new protein-coding genes that were not found in the keyword search. This approach,
438 applied to orphan crops for which few genomic resources are available, will improve the
439 annotation of their genomes.

440 Material and Methods

441 *Plant material*

442 This study used 125 genotypes from *Dioscorea alata* and two outgroup genotypes, belonging to
443 *D. rotundata* and *D. trifida* species, collected in 19 different countries. Genotypes were selected
444 to cover a maximum of worldwide diversity (Figure 1; Supplemental Table 1) based on previous
445 studies (Sharif et al., 2020). These genotypes were used to find genomic polymorphism, compared
446 to the *D. alata* reference genome. From the whole set used in this study, a total of 45 *D. alata*
447 accessions harbouring a high genetic diversity were planted from 2018 to 2020 in May each year
448 and harvested at full maturity in the following year between January and March at three different
449 locations in Guadeloupe (Godet, Roujol and Duclos). The three locations have contrasting pedo-
450 climatic conditions (Cormier et al., 2021). The phenotyping was performed on freshly harvested
451 tubers. Finally, a set of six genotypes was selected from the list, based on their previously described
452 phenotypic traits, to be used for transcriptome sequencing (Supplemental Table 1).
453
454

455 *Phenotyping for quality traits*

456 Colour was analyzed by computerized image analysis techniques (Detailed method is described in
457 Supplemental Note 1). Images were collected on three tubers per genotype (45) and location (3)
458 during the 2018 cropping season. To test the repeatability and accuracy of the system, a custom
459 colour chart was placed on every picture. Acquired images were analyzed using the Rvision library
460 (Garnier and Muschelli, 2022) with R programming language (R Core Team, 2021). Different
461 colour indexes were tested to characterize purple yam. Finally, the Hue was found to be
462 discriminant within the studied diversity panel. Brown (BI) index was also calculated according
463 to (Buera et al., 1986).

464

465 Starch content was predicted using near infrared spectroscopy (NIRS). Two replicates of yam flour
466 samples were scanned with a FOSS-NIRSystems model 6500 scanning monochromator (FOSS-
467 NIRSystems, Silver Spring, MD, USA) equipped with an autocup. The spectroscopic procedures
468 and data recording were conducted with ISIscan (TM) software (FOSS, Hillerød, Denmark). The
469 model was calibrated using 2016 and 2017 data and validated on an external independent 2018
470 dataset. (Supplemental Note 2).

471

472 The texture was measured on steam-cooked yam by penetrometry using the TAX-TPlus texture
473 analyzer (Stable Micro Systems, Ltd., Surrey, UK). Yam tubers harvested at Godet 2019 were
474 used for this analysis. Three tubers per variety were sampled and divided into three equal sections
475 (proximal, distal, central) and prepared as described in Supplemental Note 3. Each part was used
476 to produce three cubes at the central section. Each cube was steam-cooked up to 15 min, followed
477 by a cooling time of 7 min, corresponding to a cube temperature of 45°C. Total area was calculated
478 from the puncture test, while four texture profile analyzer parameters were computed from the
479 force-time curve: hardness (N), cohesiveness, gumminess, and springiness (Supplemental Note 3).

480

481 *Whole-genome sequencing*

482 The genomic DNA from 127 genotypes of *Dioscorea spp*. was extracted from leaves according to
483 a protocol using mixed alkyl trimethylammonium bromide (MATAB) buffer and NucleoMag Plant
484 Kit (Macherey–Nagel, Germany) already described by (Cormier et al., 2019). Sequencing libraries
485 were prepared as described in Dossa et al. (2023). Paired-end high throughput sequencing (2 × 150

486 bp) was performed on an Illumina NovaSeq 6000 instrument on GeT-PlaGe platform, (Toulouse,
487 France) and Genewiz company (Leipzig, Germany).
488 The reads quality and mapping and Variant discovery and filtering were performed as described
489 in Dossa et al. in press. For the filtering, the following parameters were used: --max-missing 0.5 -
490 -minQ 30 --minDP 10 --maxDP 200. For gene candidate purposes, we selected only the
491 polymorphisms annotated with a high impact on the amino acid sequences.
492
493 *Genetic population analyses*
494 Among all the 127 genotypes of *Dioscorea* spp., 107 diploid genotypes (Supplemental Table 1)
495 were selected for further population genetics analyses. We calculated the population structure
496 using the original filters described above, in addition to: --maf 0.01 --max-alleles 2 --min-alleles
497 2 --remove-indels --max-missing 0.5 --thin 10 --ld-window 50 --min-r2 0.1. The final VCF file
498 obtained from this analysis was used for population structure analysis using ADMIXTURE version
499 1.23 (Alexander and Lange, 2011). After testing the replicates of K = 1 to 15, we limited the
500 ancestry threshold to 75%, where genotypes with a value lower than this were considered in the
501 admixture group. After Bayesian clustering analysis, populations were redefined according to the
502 results obtained.
503 To identify genes under selection, *Fst* analysis was conducted using vcftools with an *Fst*-window-
504 size of 50kb and a *Fst*-window-step of 10kb. A cutoff of the 5% top values and a Weibull
505 distribution analysis on regions with extreme *Fst* between the different genetic clusters identified,
506 were used to select candidates. Subsequently, we searched for the genes on these regions using
507 BEDTools intersect version 2.29 (Quinlan and Hall, 2010).
508 The gene ontology enrichment analysis of the genes from outlier regions was performed with
509 KOBAS-i webserver (Bu et al., 2021) in order to find out their biological significance.
510
511 *Comparative genomics and orthology analyses*
512 For comparative genomics analyses 45 species from diverse plant lineages were selected
513 (Supplemental Table 2). The species used in this analysis were chosen based on the completeness
514 of their genomes, phylogenetic distribution along the tree of viridiplantae and in the use of plant-
515 models. These plants belong to the monocotyledons and dicotyledons, encompassing 17 different
516 clades. Five genomes belonging to Dioscoreales were added (*Dioscorea alata, D. rotundata, D.*

517 *dumetorum, D. zingiberensis, Trichopus zeylanicus*). The completeness score of each plant
518 proteome was verified by BUSCO (Simão et al., 2015) using the Viridiplantae dataset of BUSCO.
519 We chose a threshold of 70% of completeness to include public proteomes. For the inference of
520 orthologous groups, OrthoFinder software v2.4.0 (Emms and Kelly, 2015) was used with the
521 default parameters. Gene presence and absence of species in orthogroups was represented by
522 UpSetR (Lex et al., 2014), using a binary matrix as input.
523
524 *Metabolic pathways selection and gene retrieval from D. alata*
525 We downloaded the complete list of enzyme codes (EC) from the pathway of pentose and
526 glucuronate interconversions (map00040), starch and sucrose metabolism (map00500) and
527 flavonoid biosynthesis (map00941) from KEGG database (https://www.genome.jp). These EC
528 numbers were used to retrieve the genes annotated in the GFF file of *D. alata* (https://phytozome-
529 next.jgi.doe.gov/info/Dalata_v2_1). The genes annotated with these EC numbers were further
530 searched in the orthogroups produced by our comparative genomics study, and all the homologous
531 genes from *D. alata* were retrieved. The VCF file was reduced to the middle coding sequences of
532 the annotated genes of interest.
533
534 *Candidate gene association analysis*
535 Three VCF files, one for each target metabolic pathway genes, containing non-synonymous SNPs
536 were extracted from the whole VCF file and further filtered for minor allele frequency ≥ 0.05,
537 missing rate $< 20\%$ using TASSEL5.0 (Bradbury et al., 2007). The associations between the
538 polymorphisms from each metabolic pathway candidate genes and the corresponding phenotypic
539 traits were tested using TASSEL5.0 based on the General Linear Model. Only significant variants
540 with $P \leq 0.001$ were retained. The association analysis was conducted with phenotypic data from
541 each location independently. The effect of alleles at significant SNPs was assessed by comparing
542 phenotyping data for haplotype groups. A Student's t-test was used to compare the groups of
543 haplotypes ($P<0.05$) in the R4.0.23 software with the "ggpubr" packages and "rstatix"
544
545 *Transcriptome analysis of six D. alata genotypes*
546 RNA was extracted from six of the most *D. alata* diverse genotypes, based on phenotypic traits.
547 Three biological replicates for each genotype were used. For each genotype, a 50 mg sample cut

548 from the middle of freshly harvested tuber was ground in liquid nitrogen and total cellular RNA
549 was extracted using a Sigma-Aldrich (St. Louis, MO) Spectrum™ Plant Total RNA kit with a
550 DNAse treatment. This was followed by quantification by the Invitrogen (Carlsbad, CA) Quant-
551 iT™ RiboGreen® RNA Reagent based on the manufacturer's protocol and verification of RNA
552 quality by 5200 Fragment Analyzer™ System (Agilent, Santa Clara, CA) profile.
553 Synthesis of cDNA and construction of libraries were done with Illumina, Inc. TruSeq RNA
554 Sample Preparation v2 Kit (San Diego, CA). Fragment size of selected cDNA were between 200pb
555 and 400pb. The 18 libraries were indexed, mixed and sequenced using one lane of an Illumina
556 HiSeq 3000 sequencer with the 2 x 150 cycles, paired-end, indexed protocol (Genewiz facility,
557 Liepzig, Germany).
558 The quality of the raw reads from the 18 libraries were assessed by FastQC version 0.11.7
559 (Andrews, 2010). The low-quality sequences and the Illumina adapters were trimmed by
560 Trimmomatic version 0.39 (Bolger et al., 2017). Trimmed reads were quantified using Kallisto
561 version 0.46.1 (Bray et al., 2016) using the reference genome of *D. alata* (Bredeson et al., 2022).
562 To conduct the expression analysis of the genes, we used the log10 of the average value of
563 transcripts per million over the three biological replicates. A functional annotation was performed
564 using the conserved domains of PFAM version 35 and a gene ontology annotation using Pfam2go
565 (http://current.geneontology.org/ontology/external2go/pfam2go).
566 All the bioinformatics analyses were performed with the support of MESO@LR-Platform at
567 University of Montpellier, CIRAD UMR-AGAP HPC (South Green Platform) and IFB core
568 cluster.

569 DATA AVAILABILITY

570 The Illumina HiSeq 3000 sequencing raw data are available in the NCBI SRA (Sequence Read
571 Archive), and transcriptome data are under the BioProject number: PRJNA918625. The
572 phenotypic datasets are available from the corresponding author upon request.

573

FUNDING

We acknowledge the support from Breeding RTB Products for End User Preferences (RTBfoods) Project (Grant OPP1178942) through funds received from the Bill and Melinda Gates Foundation. The production of the genomic resources was supported by a fund from the CGIAR Research Program on Roots, Tubers and Bananas (CRP-RTB).

ACKNOWLEDGMENTS

We are grateful to Christian Mestres for his valuable advice on yam quality and its attributes. We gratefully acknowledge Nancy Terrier for her critical reading of the manuscript. Our greatest debts go to CRP-RTB, IITA and PRC genebanks and Juliane Kaoh, Mamy Tiana Rajaonah, Senanayake Ravinda Lakshan, and Babil Pachakkil for providing us with plant leaves material. We are also thankful to Elie Nudol, Marie-Claire Gravillon, Christophe Perrot and Erick Maledon for their invaluable help with field work and phenotyping, and Sandrine Causse and Hélène Vignes for laboratory assistance in preparing *D. alata* libraries. Finally, special thanks are due to Sylvain Santoni and Muriel Latreille for their assistance in transcriptome preparation.

589 # REFERENCES

590

591 **Alexander, D. H., and Lange, K.** (2011). Enhancements to the ADMIXTURE algorithm for
592 individual ancestry estimation. *BMC Bioinformatics* **12**:246.

593 **Alves-Pereira, A., Zucchi, M. I., Clement, C. R., Viana, J. P. G., Pinheiro, J. B., Veasey, E.**
594 **A., and de Souza, A. P.** (2022). Selective signatures and high genome-wide diversity in
595 traditional Brazilian manioc (Manihot esculenta Crantz) varieties. *Sci. Rep.* **12**:1268.

596 **Andrews, S. and others** (2010). FastQC: a quality control tool for high throughput sequence
597 data Advance Access published 2010.

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

601 **Atmodjo, M. A., Hao, Z., and Mohnen, D.** (2013). Evolving Views of Pectin Biosynthesis.
602 *Annu. Rev. Plant Biol.* **64**:747–779.

603 **Banda, L., Kyallo, M., Domelevo Entfellner, J.-B., Moyo, M., Swanckaert, J., Mwanga, R.**
604 **O. M., Onyango, A., Magiri, E., Gemenet, D. C., Yao, N., et al.** (2021). Analysis of β-
605 amylase gene (Amyβ) variation reveals allele association with low enzyme activity and
606 increased firmness in cooked sweetpotato (Ipomoea batatas) from East Africa. *J. Agric.*
607 *Food Res.* **4**:100121.

608 **Beleia, A., Butarelo, S. S., and Silva, R. S. F.** (2006). Modeling of starch gelatinization during
609 cooking of cassava (Manihot esculenta Crantz). *LWT - Food Sci. Technol.* **39**:400–405.

610 **Bhosale, S. H., Rao, M. B., and Deshpande, V. V.** (1996). Molecular and industrial aspects of
611 glucose isomerase. *MICROBIOL REV* **60**.

612 **Bolger, M. E., Arsova, B., and Usadel, B.** (2017). Plant genome and transcriptome annotations:
613 from misconceptions to simple solutions. *Brief. Bioinform.* Advance Access published
614 January 5, 2017, doi:10.1093/bib/bbw135.

615 **Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E.**
616 **S.** (2007). TASSEL: software for association mapping of complex traits in diverse
617 samples. *Bioinformatics* **23**:2633–2635.

618 **Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L.** (2016). Near-optimal probabilistic
619 RNA-seq quantification. *Nat. Biotechnol.* **34**:525–527.

620 **Bredeson, J. V., Lyons, J. B., Oniyinde, I. O., Okereke, N. R., Kolade, O., Nnabue, I.,**
621 **Nwadili, C. O., Hřibová, E., Parker, M., Nwogha, J., et al.** (2022). Chromosome
622 evolution and the genetic basis of agronomically important traits in greater yam. *Nat.*
623 *Commun.* **13**:2001.

624 **Bu, D., Luo, H., Huo, P., Wang, Z., Zhang, S., He, Z., Wu, Y., Zhao, L., Liu, J., Guo, J., Fang,**
625 **S., Cao, W., Yi, L., Zhao, Y., & Kong, L.** (2021). KOBAS-i: intelligent prioritization and
626 exploratory visualization of biological functions for gene enrichment analysis. *Nucleic*
627 *Acids Research* **49**:W317–W325.

628 **Buera, M. P., Lozano, R., and Petriella, C.** (1986). Definition of colour in the non enzymatic
629 browning process. *Farbe* **32**:318–322.

630 **Burton, R. A., and Fincher, G. B.** (2012). Current challenges in cell wall biology in the cereals
631 and grasses. *Front. Plant Sci.* **3**.

632 **Butelli, E., Licciardello, C., Ramadugu, C., Durand-Hulak, M., Celant, A., Reforgiato**
633 **Recupero, G., Froelicher, Y., and Martin, C.** (2019). Noemi Controls Production of
634 Flavonoid Pigments and Fruit Acidity and Illustrates the Domestication Routes of
635 Modern Citrus Varieties. *Curr. Biol.* **29**:158-164.e2.

636 **Chair, H., Arnau, G., and Mota, A. P. Z.** Yam Genomics. In *Underutilised Crop Genomes*, pp.
637 373–398. Springer International Publishing.

638 **Cormier, F., Lawac, F., Maledon, E., Gravillon, M.-C., Nudol, E., Mournet, P., Vignes, H.,**
639 **Chaïr, H., and Arnau, G.** (2019). A reference high-density genetic map of greater yam
640 (Dioscorea alata L.). *TAG Theor. Appl. Genet. Theor. Angew. Genet.* **132**:1733.

641 **Cormier, F., Martin, G., Vignes, H., Lachman, L., Cornet, D., Faure, Y., Maledon, E.,**
642 **Mournet, P., Arnau, G., and Chaïr, H.** (2021). Genetic control of flowering in greater
643 yam (Dioscorea alata L.). *BMC Plant Biol.* **21**:163.

644 **Critchley, J. H., Zeeman, S. C., Takaha, T., Smith, A. M., and Smith, S. M.** (2001). A critical
645 role for disproportionating enzyme in starch breakdown is revealed by a knock-out
646 mutation in Arabidopsis: Function of D-enzyme in starch breakdown. *Plant J.* **26**:89–100.

647 **Das, M., Haberer, G., Panda, A., Das Laha, S., Ghosh, T. C., and Schäffner, A. R.** (2016).
648 Expression Pattern Similarities Support the Prediction of Orthologs Retaining Common
649 Functions after Gene Duplication Events. *Plant Physiol.* **171**:2343–2357.

650 **do Nascimento, J. R. O., Júnior, A. V., Bassinello, P. Z., Cordenunsi, B. R., Mainardi, J. A.,**
651 **Purgatto, E., and Lajolo, F. M.** (2006). Beta-amylase expression and starch degradation
652 during banana ripening. *Postharvest Biol. Technol.* **40**:41–47.

653 **Domínguez, M., Dugas, E., Benchouaia, M., Leduque, B., Jiménez-Gómez, J. M., Colot, V.,**
654 **and Quadrana, L.** (2020). The impact of transposable elements on tomato diversity. *Nat.*
655 *Commun.* **11**:4058.

656 **Dong, M.-Y., Fan, X.-W., and Li, Y.-Z.** (2019). Cassava AGPase genes and their encoded
657 proteins are different from those of other plants. *Planta* **250**:1621–1635.

658 **Dossa, K., Morel, A., Houngbo, E.M., Mota, A.Z., Maledon, E. et al.** (2023). Genome-wide
659 association studies reveal novel loci controlling tuber flesh colour and oxidative browning
660 in *Dioscorea alata*. *bioRxiv* 2023.03.12.532275.

661 **Dufour, D., Hershey, C., Hamaker, B. R., and Lorenzen, J.** (2021). Integrating end-user
662 preferences into breeding programmes for roots, tubers and bananas. *Int. J. Food Sci.*
663 *Technol.* **56**:1071–1075.

664 **Ehounou, A. E., Cormier, F., Maledon, E., Nudol, E., Vignes, H., Gravillon, M. C.,**
665 **N'guetta, A. S. P., Mournet, P., Chaïr, H., Kouakou, A. M., et al.** (2022).
666 Identification and validation of QTLs for tuber quality related traits in greater yam
667 (*Dioscorea alata* L.). *Sci. Rep.* **12**:8423.

668 **Emms, D. M., and Kelly, S.** (2015). OrthoFinder: solving fundamental biases in whole genome
669 comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.*
670 **16**:157.

671 **Falcone Ferreyra, M. L., Rius, S. P., and Casati, P.** (2012). Flavonoids: biosynthesis,
672 biological functions, and biotechnological applications. *Front. Plant Sci.* **3**.

673 **Fang, H., Zhang, F., Zhang, C., Wang, D., Shen, S., He, F., Tao, H., Wang, R., Wang, M.,**
674 **Wang, D., et al.** (2022). Function of hydroxycinnamoyl transferases for the biosynthesis
675 of phenolamides in rice resistance to Magnaporthe oryzae. *J. Genet. Genomics* **49**:776–
676 786.

677 **Garcia-Gimenez, G., Barakate, A., Smith, P., Stephens, J., Khor, S.F., Doblin, M.S., Hao,**
678 **P., Bacic, A., Fincher, G.B., Burton, R.A., Waugh, R., Tucker, M.R. and Houston,**
679 **K.** (2020), Targeted mutation of barley (1,3;1,4)-β-glucan synthases reveals complex
680 relationships between the storage and cell wall polysaccharide content. Plant J., 104:
681 1009-1022.

682 **Garnier, S., and Muschelli, J.** (2022). *Rvision - A computer vision library for R*.

683 **Gatarira, C., Agre, P., Matsumoto, R., Edemodu, A., Adetimirin, V., Bhattacharjee, R.,**
684 **Asiedu, R., and Asfaw, A.** (2020). Genome-Wide Association Analysis for Tuber Dry
685 Matter and Oxidative Browning in Water Yam (Dioscorea alata L.). *Plants* **9**:969.

686 **Girma, G., Hyma, K. E., Asiedu, R., Mitchell, S. E., Gedil, M., and Spillane, C.** (2014).
687 Next-generation sequencing based genotyping, cytometry and phenotyping for
688 understanding diversity and evolution of guinea yams. *Theor. Appl. Genet.* **127**:1783–
689 1794.

690 **Hardigan, M. A., Laimbeer, F. P. E., Newton, L., Crisovan, E., Hamilton, J. P.,**
691 **Vaillancourt, B., Wiegert-Rininger, K., Wood, J. C., Douches, D. S., Farré, E. M., et**
692 **al.** (2017). Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary
693 history and targets of domestication in the cultivated potato. *Proc. Natl. Acad. Sci.* **114**.

694 **He, Y., Zhu, D., Sun, Y., Wang, Q., Zhu, L., and Zeng, H.** (2021). Metabonomic Profiling
695 Analyses Reveal ANS Upregulation to Enhance the Flavonoid Pathway of Purple-
696 Fleshed Sweet Potato Storage Root in Response to Deep Shading. *Agronomy* **11**:737.

697 **Hendriks, J. H. M., Kolbe, A., Gibon, Y., Stitt, M., and Geigenberger, P.** (2003). ADP-
698 Glucose Pyrophosphorylase Is Activated by Posttranslational Redox-Modification in
699 Response to Light and to Sugars in Leaves of Arabidopsis and Other Plant Species. *Plant*
700 *Physiol.* **133**:838–849.

701 **Honfozo, L., Adinsi, L., Bouniol, A., Adetonah, S., Forsythe, L., Kleih, U., Hounhouigan, J.**
702 **D., Fliedel, G., and Akissoe, N. H.** (2021). Boiled yam end-user preferences and
703 implications for trait evaluation. *Int. J. Food Sci. Technol.* **56**:1447–1457.

704 **Hou, J., Liu, T., Reid, S., Zhang, H., Peng, X., Sun, K., Du, J., Sonnewald, U., and Song, B.**
705 (2019). Silencing of α-amylase StAmy23 in potato tuber leads to delayed sprouting. *Plant*
706 *Physiol. Biochem.* **139**:411–418.

707 **Janeček, Š., Svensson, B., and MacGregor, E. A.** (2014). α-Amylase: an enzyme specificity
708 found in various families of glycoside hydrolases. *Cell. Mol. Life Sci.* **71**:1149–1170.

709 **Jiang, F., Wang, J.-Y., Jia, H.-F., Jia, W.-S., Wang, H.-Q., and Xiao, M.** (2013). RNAi-
710 Mediated Silencing of the Flavanone 3-Hydroxylase Gene and Its Effect on Flavonoid
711 Biosynthesis in Strawberry Fruit. *J. Plant Growth Regul.* **32**:182–190.

712 **Keeling, P. L., and Myers, A. M.** (2010). Biochemistry and Genetics of Starch Synthesis. *Annu.*
713 *Rev. Food Sci. Technol.* **1**:271–303.

714 **Kim, Y.-L., Mun, S., Park, K.-H., Shim, J.-Y., and Kim, Y.-R.** (2013). Physicochemical
715 functionality of 4-α-glucanotransferase-treated rice flour in food application. *Int. J. Biol.*
716 *Macromol.* **60**:422–426.

717 **Kolbe, A., Tiessen, A., Schluepmann, H., Paul, M., Ulrich, S., and Geigenberger, P.** (2005).
718 Trehalose 6-phosphate regulates starch synthesis via posttranslational redox activation of
719 ADP-glucose pyrophosphorylase. *Proc. Natl. Acad. Sci.* **102**:11118–11123.

720 **Kouadio, O. K., N'dri, D. Y., Nindjin, C., Marti, A., Casiraghi, M. C., Faoro, F., Erba, D.,**
721 **Bonfoh, B., and Amani, N. G.** (2013). Effect of resistant starch on the cooking quality of
722 yam (*Dioscorea* spp.) and cassava (*Manihot esculenta*) based paste products. *Int. J.*
723 *Food Sci. Nutr.* **64**:484–493.

724 **Leah, R., Kigel, J., Svendsen, I., and Mundy, J.** (1995). Biochemical and Molecular
725 Characterization of a Barley Seed β-Glucosidase. *J. Biol. Chem.* **270**:15789–15797.

726 **Lebot, V., Trilles, B., Noyer, J. L., and Modesto, J.** Genetic relationships between Dioscorea
727 alata L. cultivars.

728 **Leng, Y., Yang, Y., Ren, D., Huang, L., Dai, L., Wang, Y., Chen, L., Tu, Z., Gao, Y., Li, X.,**
729 **et al.** (2017). A Rice *PECTATE LYASE-LIKE* Gene Is Required for Plant Growth and
730 Leaf Senescence. *Plant Physiol.* **174**:1151–1166.

731 **Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H.** (2014). UpSet:
732 Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.* **20**:1983–1992.

733 **Li, X., Shin, S., Heinen, S., Dill-Macky, R., Berthiller, F., Nersesian, N., Clemente, T.,**
734 **McCormick, S., and Muehlbauer, G. J.** (2015). Transgenic Wheat Expressing a Barley
735 UDP-Glucosyltransferase Detoxifies Deoxynivalenol and Provides High Levels of
736 Resistance to *Fusarium graminearum*. *Mol. Plant-Microbe Interactions®* **28**:1237–1246.

737 **Li, C., Wu, H.-M., and Cheung, A. Y.** (2016). FERONIA and Her Pals: Functions and
738 Mechanisms. *Plant Physiol.* **171**:2379–2392.

739 **Li, X.J., Zhang, J.Q., Wu, Z.C., Lai, B., Huang, X.M., Qin, Y.H., Wang, H.C., Hu, G.B.**
740 (2016). Functional characterization of a glucosyltransferase gene, LcUFGT1, involved in
741 the formation of cyanidin glucoside in the pericarp of *Litchi chinensis*. *Physiol. Plant*.
742 **156**:139-149.
743

744 **Liao, L., Zhang, W., Zhang, B., Fang, T., Wang, X.-F., Cai, Y., Ogutu, C., Gao, L., Chen,**
745 **G., Nie, X., et al.** (2021). Unraveling a genetic roadmap for improved taste in the
746 domesticated apple. *Mol. Plant* **14**:1454–1471.

747 **Liu, C.; Yu, Q.; Li, Z.; Jin, X.; Xing, W.** (2020). Metabolic and transcriptomic analysis related
748 to flavonoid biosynthesis during the colour formation of *Michelia crassipes* tepal. *Plant*
749 *Physiology and Biochemistry* **155**: 938-951.

750 **Lou, H., Tucker, M. R., Shirley, N. J., Lahnstein, J., Yang, X., Ma, C., Schwerdt, J., Fusi,**
751 **R., Burton, R. A., Band, L. R., et al.** (2022). The *cellulose synthase-like F3* (*CslF3*)
752 gene mediates cell wall polysaccharide synthesis and affects root growth and
753 differentiation in barley. *Plant J.* **110**:1681–1699.

754 **Marks, R. A., Hotaling, S., Frandsen, P. B., and VanBuren, R.** (2021). Representation and
755 participation across 20 years of plant genome sequencing. *Nat. Plants* **7**:1571–1578.

756 **Meyer, R. S., and Purugganan, M. D.** (2013). Evolution of crop species: genetics of
757 domestication and diversification. *Nat. Rev. Genet.* **14**:840–852.

758 **Miao, H., Sun, P., Liu, Q., Jia, C., Liu, J., Hu, W., Jin, Z., and Xu, B.** (2017). Soluble Starch
759 Synthase III-1 in Amylopectin Metabolism of Banana Fruit: Characterization,
760 Expression, Enzyme Activity, and Functional Analyses. *Front. Plant Sci.* **8**.

761 **Mota, A. P. Z., Fernandez, D., Arraes, F. B. M., Petitot, A.-S., de Melo, B. P., de Sa, M. E.**
762 **L., Grynberg, P., Saraiva, M. A. P., Guimaraes, P. M., Brasileiro, A. C. M., et al.**
763 (2020). Evolutionarily conserved plant genes responsive to root-knot nematodes
764 identified by comparative genomics. *Mol. Genet. Genomics* **295**:1063–1078.

765 **Mu, W., Hassanin, H. A. M., Zhou, L., and Jiang, B.** (2018). Chemistry Behind Rare Sugars
766 and Bioprocessing. *J. Agric. Food Chem.* **66**:13343–13345.

767 **Muñoz, F. J., Baroja-Fernández, E., Morán-Zorzano, M. T., Alonso-Casajús, N., and**
768 **Pozueta-Romero, J.** (2006). Cloning, Expression and Characterization of a Nudix
769 Hydrolase that Catalyzes the Hydrolytic Breakdown of ADP-glucose Linked to Starch
770 Biosynthesis in Arabidopsis thaliana. *Plant Cell Physiol.* **47**:926–934.

771 **Ngolong Ngea, G. L., Guillon, F., Essia Ngang, J. J., Bonnin, E., Bouchet, B., and Saulnier,**
772 **L.** (2016). Modification of cell wall polysaccharides during retting of cassava roots. *Food*
773 *Chem.* **213**:402–409.

774 **Orkwor, G. C., Asiedu, R., and Ekanayake, I.** (1998). *Food yams: Advances in research*.

775 **Pont, C., Murat, F., Confolent, C., Balzergue, S., and Salse, J.** (2011). RNA-seq in grain
776 unveils fate of neo- and paleopolyploidization events in bread wheat (Triticum aestivum
777 L.). *Genome Biol.* **12**:R119.

778 **Prüfer, K., Muetzel, B., Do, H.-H., Weiss, G., Khaitovich, P., Rahm, E., Pääbo, S.,**
779 **Lachmann, M., and Enard, W.** (2007). FUNC: a package for detecting significant
780 associations between gene sets and ontological annotations. *BMC Bioinformatics* **8**:41.

781 **Quinlan, A. R., and Hall, I. M.** (2010). BEDTools: a flexible suite of utilities for comparing
782 genomic features. *Bioinformatics* **26**:841–842.

783 **Richmond, T. A., and Somerville, C. R.** (2000). The Cellulose Synthase Superfamily. *Plant*
784 *Physiol.* **124**:495–498.

785 **Ross, H. A., Morris, W. L., Ducreux, L. J. M., Hancock, R. D., Verrall, S. R., Morris, J. A.,**
786 **Tucker, G. A., Stewart, D., Hedley, P. E., McDougall, G. J., et al.** (2011). Pectin
787 engineering to modify product quality in potato: Pectin engineering in potato. *Plant*
788 *Biotechnol. J.* **9**:848–856.

789 **Sampedro, J., Valdivia, E. R., Fraga, P., Iglesias, N., Revilla, G., and Zarra, I.** (2017).
790 Soluble and Membrane-Bound β-Glucosidases Are Involved in Trimming the Xyloglucan
791 Backbone. *Plant Physiol.* **173**:1017–1030.

792 **Sharif, B. M., Burgarella, C., Cormier, F., Mournet, P., Causse, S., Van, K. N., Kaoh, J.,**
793 **Rajaonah, M. T., Lakshan, S. R., Waki, J., et al.** (2020). Genome-wide genotyping
794 elucidates the geographical diversification and dispersal of the polyploid and clonally
795 propagated yam (*Dioscorea alata*). *Ann. Bot.* **126**:1029–1038.

796 **Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M.**
797 (2015). BUSCO: assessing genome assembly and annotation completeness with single-
798 copy orthologs. *Bioinformatics* **31**:3210–3212.

799 **Singh, K., Rani, A., Kumar, S., Sood, P., Mahajan, M., Yadav, S. K., Singh, B., and Ahuja,**
800 **P. S.** (2008). An early gene of the flavonoid pathway, flavanone 3-hydroxylase, exhibits a
801 positive relationship with the concentration of catechins in tea (*Camellia sinensis*). *Tree*
802 *Physiol.* **28**:1349–1356.

803 **Siqueira, M. V., Dequigiovanni, G., Corazon-Guivin, M. A., Feltran, J. C., and Veasey, E.**
804 **A.** (2012). DNA fingerprinting of water yam (*Dioscorea alata*) cultivars in Brazil based
805 on microsatellite markers. *Hortic. Bras.* **30**:653–659.

806 **Soliman, A., Ayele, B. T., and Daayf, F.** (2014). Biochemical and Molecular Characterization
807 of Barley Plastidial ADP-Glucose Transporter (HvBT1). *PLoS ONE* **9**:e98524.

808 **Taylor, N. G.** (2008). Cellulose biosynthesis and deposition in higher plants. *New Phytol.*
809 **178**:239–252.

810 **Tetlow, I. J.** (2011). Starch biosynthesis in developing seeds. *Seed Sci. Res.* **21**:5–32.

811 **Van Harsselaar, J. K., Lorenz, J., Senning, M., Sonnewald, U., and Sonnewald, S.** (2017).
812 Genome-wide analysis of starch metabolism genes in potato (*Solanum tuberosum* L.).
813 *BMC Genomics* **18**:37.

814 **Watcharatpong, P., Kaga, A., Chen, X., and Somta, P.** (2020) Narrowing Down a Major QTL
815 Region Conferring Pod Fiber Contents in Yardlong Bean (*Vigna unguiculata*), a
816 Vegetable Cowpea. *Genes*, **11**, 363.

817 **Wiberley-Bradford, A. E., Busse, J. S., Jiang, J., and Bethke, P. C.** (2014). Sugar
818 metabolism, chip colour, invertase activity, and gene expression during long-term cold
819 storage of potato (*Solanum tuberosum*) tubers from wild-type and vacuolar invertase
820 silencing lines of Katahdin. *BMC Res. Notes* **7**:801.

821 **Wu, Z. G., Li, X. X., Lin, X. C., Jiang, W., Tao, Z. M., Mantri, N., Fan, C. Y., and Bao, X.**
822 **Q.** (2014). Genetic diversity analysis of yams (Dioscorea spp.) cultivated in China using
823 ISSR and SRAP markers. *Genet. Resour. Crop Evol.* **61**:639–650.

824 **Wu, Z.-G., Jiang, W., Mantri, N., Bao, X.-Q., Chen, S.-L., and Tao, Z.-M.** (2015).
825 Transciptome analysis reveals flavonoid biosynthesis regulation and simple sequence
826 repeats in yam (*Dioscorea alata* L.) tubers. *BMC Genomics* **16**:346.

827 **Xia, E., Tong, W., Hou, Y., An, Y., Chen, L., Wu, Q., Liu, Y., Yu, J., Li, F., Li, R., et al.**
828 (2020). The Reference Genome of Tea Plant and Resequencing of 81 Diverse Accessions
829 Provide Insights into Its Genome Evolution and Adaptation. *Mol. Plant* **13**:1013–1026.

830 **Xiao, N., Pan, C., Li, Y., Wu, Y., Cai, Y., Lu, Y., Wang, R., Yu, L., Shi, W., Kang, H., et al.**
831 (2021). Genomic insight into balancing high yield, good quality, and blast resistance of
832 japonica rice. *Genome Biol.* **22**:283.

833 **Zhang, J., and Sun, X.** (2021). Recent advances in polyphenol oxidase-mediated plant stress
834 responses. *Phytochemistry* **181**:112588.

835 **Zhou, Y.-X., Chen, Y.-X., Tao, X., Cheng, X.-J., and Wang, H.-Y.** (2016). Isolation and
836 characterization of cDNAs and genomic DNAs encoding ADP-glucose
837 pyrophosphorylase large and small subunits from sweet potato. *Mol. Genet. Genomics*
838 **291**:609–620.

839 **Zuker, A., Tzfira, T., Ben-Meir, H., Ovadis, M., Shklarman, E., Forkmann, G., Martens, S.,**
840 **Neta-Sharir, I., Weiss, D., and Vainstein, A.** (2002) Modification of flower colour and
841 fragrance by antisense suppression of the flavanone 3-hydroxylase gene. *Molecular*
842 *Breeding* **9** :33–41

843
844

845 # Figure Legends

846

847 **Figure 1: Population structure of *Dioscorea alata*.** A: Geographical origin of the 125 *Dioscorea*
848 *alata* accessions. B: Circos plot of the 20 chromosomes of *D. alata*, the circles represent from
849 outer to the inner circle: chromosomes, gene density, localization of genes from pentose and
850 glucuronate interconversions, genes from starch and sucrose metabolism and genes from flavonoid
851 biosynthesis. C: Principal component analysis depicting the 107 diploid genotypes of *D. alata*.
852 Accessions are coloured according to their assignment to the three genetic clusters after Admixture
853 analysis. The threshold to assign a genotype to a cluster is set at 75 %. D: Admixture barplot
854 showing the distribution of the $K = 2$ to $K = 4$ genetic clusters.

855

856 **Figure 2: Selection signature found from the comparison of clusters.** A: Manhattan plot of Fst
857 analysis of cluster A *versus* B, each colour represents one chromosome. B: *Fst* analysis of clusters
858 A *versus* C. C: *Fst* analysis of clusters B *versus* C. D: Venn diagram of the common genes found
859 among the three clusters, and the number of common outlier regions. E: The gene ontology
860 enrichment of genes found in outlier regions against the whole set of genes.

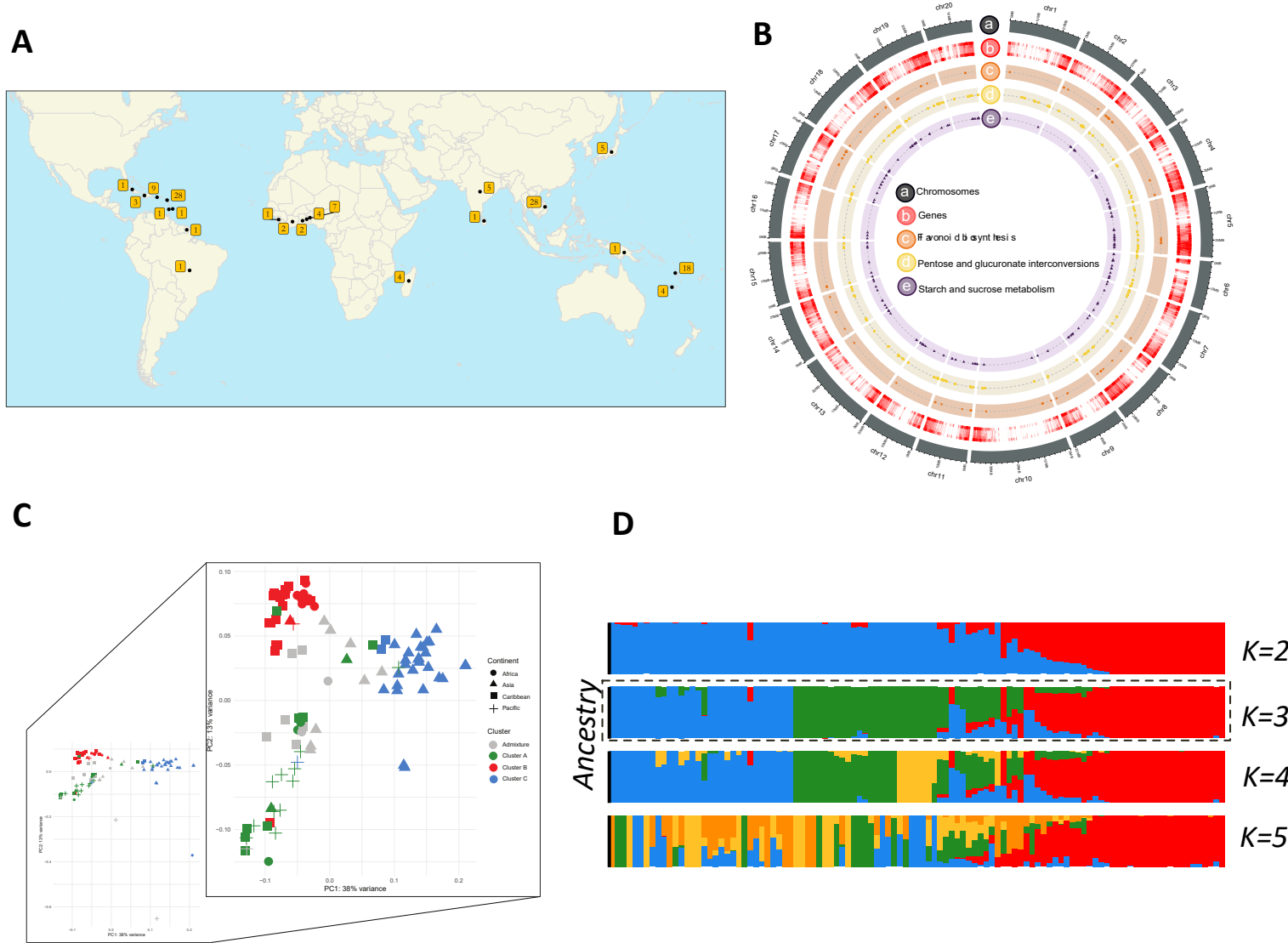
861

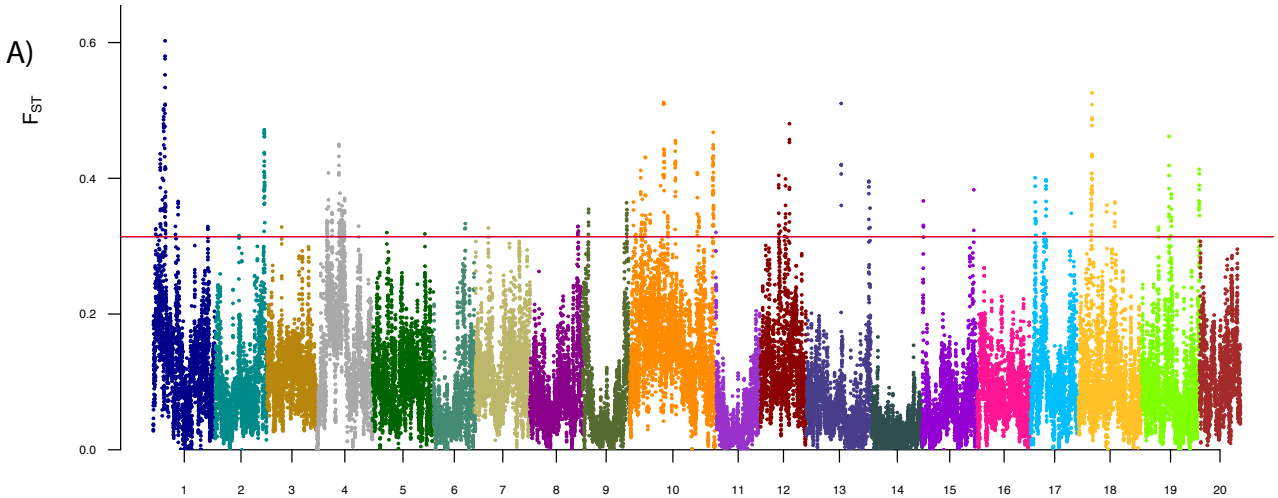
862 **Figure 3: Orthology analysis of 45 plant species.** A: Number of genes found in the keyword
863 analysis per pathway and the EC number which they were associated with (light grey). Number of
864 genes found in the orthology analysis, number of EC numbers and number of orthologous groups
865 found (dark grey). B: UpSetR of the comparison of all the monocotyledon plants used in orthology
866 analysis. Each dot represents a connection, and the bar represents the number of orthologous
867 groups found for each association. The red dots show the species from the Dioscoreales family,
868 and the red numbers represent the exclusive associations of Dioscoreales. C: UpSetR of the five
869 Dioscoreales species. D: Intersection between the genes found in the *Fst* analysis and in the
870 pathways.

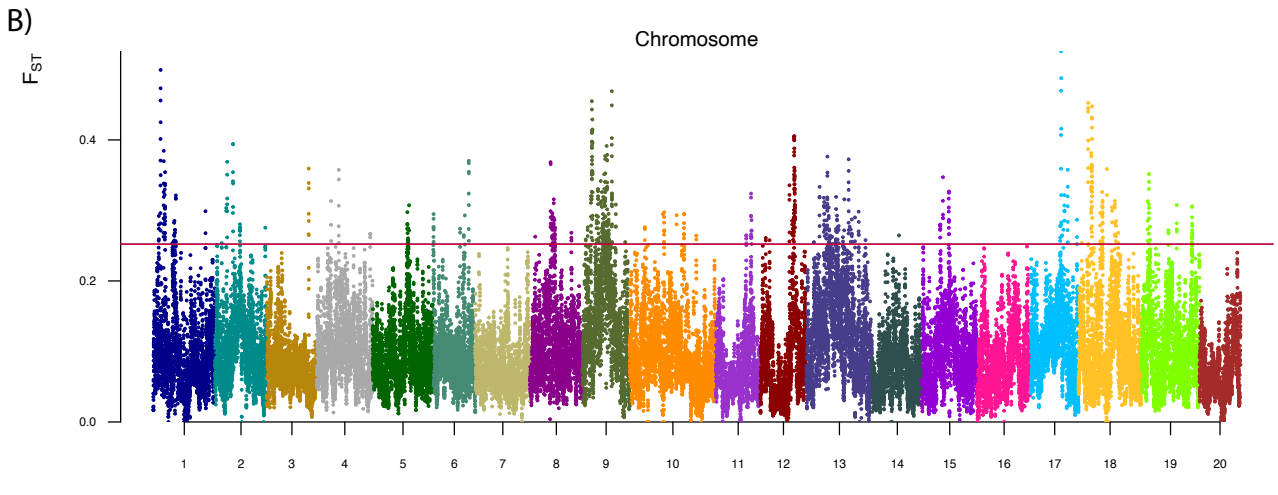
871

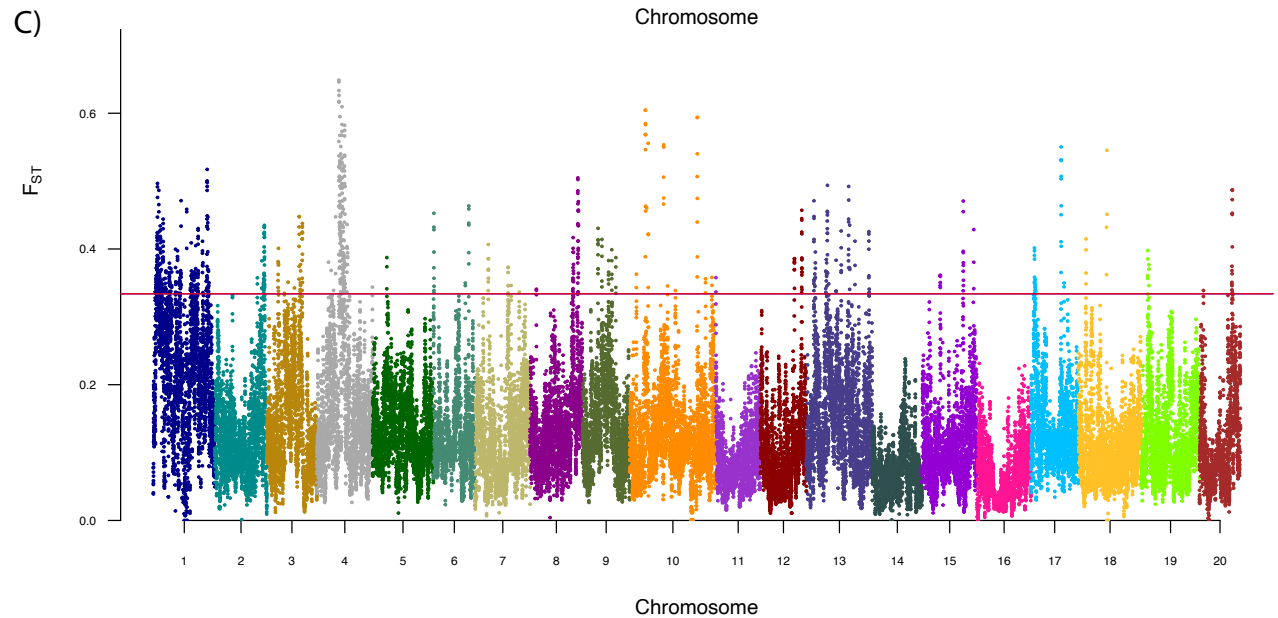
872 **Figure 4: Candidate genes associations and gene expression profile.** Selected genes and their
873 expression profile, and allelic effect on colour and texture traits. A: gene structures
874 (*Dioal.19G083800, Dioal.04G073400* and *Dioal.04G130600*) with positions of the significant
875 SNPs, their alleles and resulting amino acid changes; S4_20715506 (C/T) effect on B: gumminess;
876 C: total area; S10_17029453 (G/C) effect on D: brown index (BI); E: Transcriptome based
877 expression of *Dioal.19G083800* between two genotypes with G and C alleles (FPKM= fragments
878 per kilobase of transcript per million fragments mapped); FPKM values was extracted from three
879 biological replicates. S4_16031453 (C/T) effect on F: Hue index leading purple colour change
880 displayed by the two representative tubers. Means were separated by two-tailed *t*-test at 0.05
881 probability.

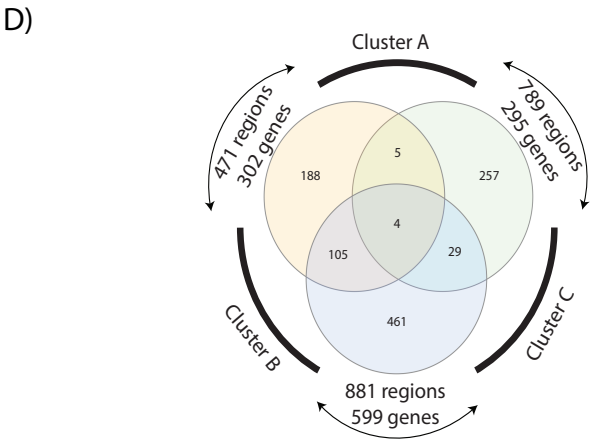
882


A
B
C
D
Chromosomes
Genes
Pentose and glucuronate interconversions
Starch and sucrose metabolism
Continent
Africa
Asia
Caribbean
Pacific
Cluster
Admixture
Cluster A
Cluster B
Cluster C
PC1: 38% variance
PC2: 13% variance
Ancestry
K=2
K=3
K=4
K=5

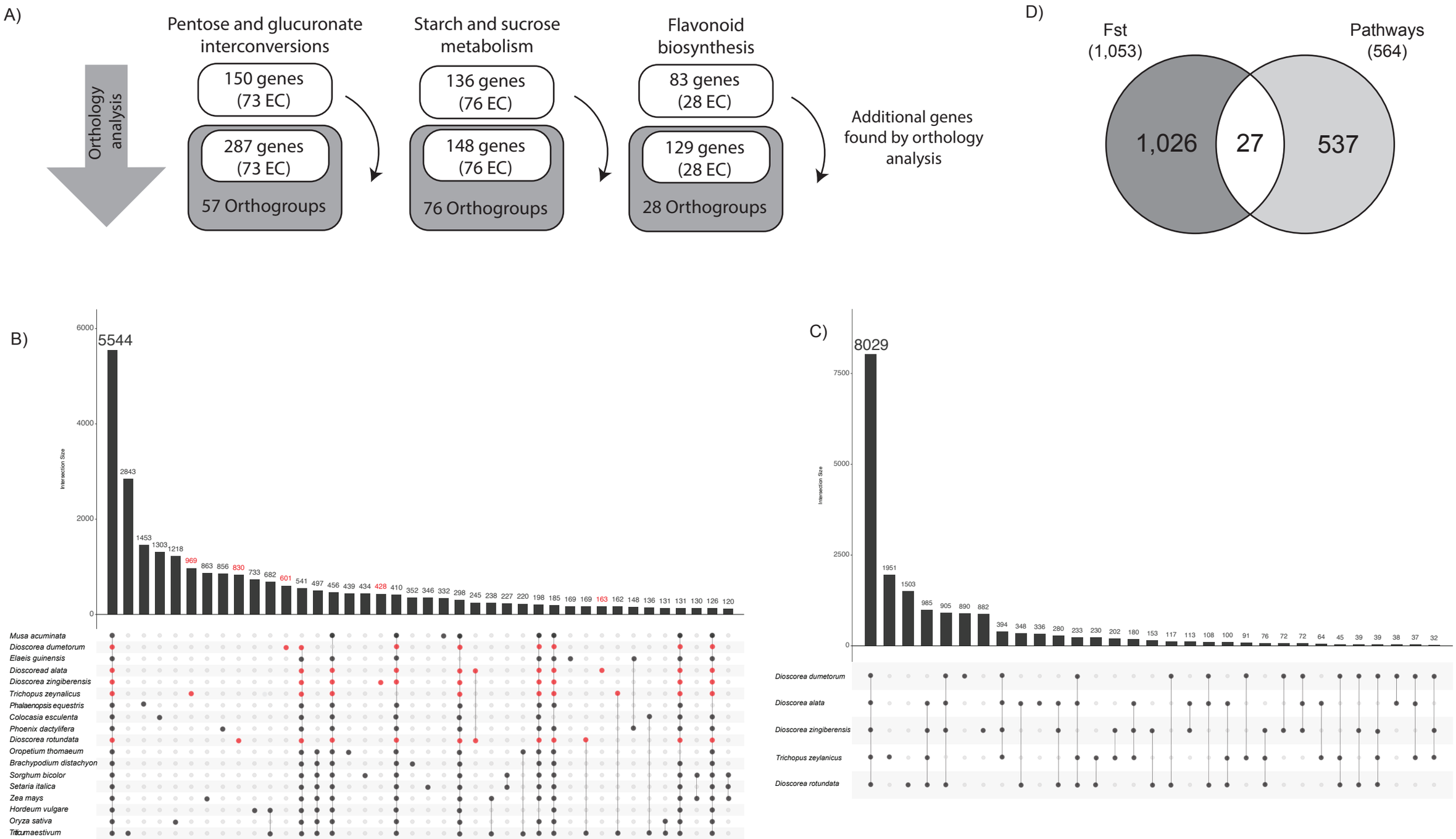

A)
F_{ST}
0.0
0.2
0.4
0.6
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Chromosome

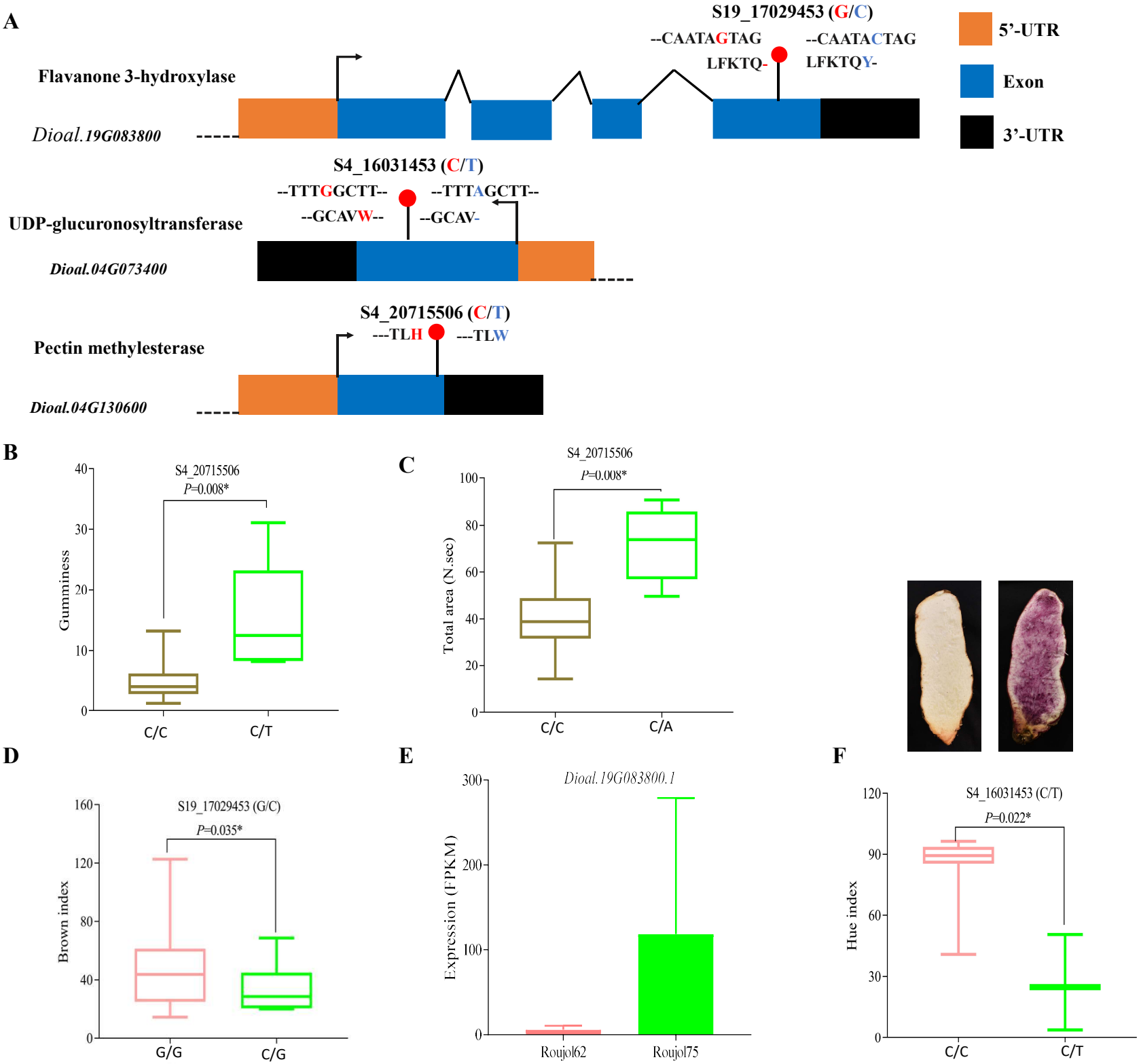

B)
F_{ST}
0.0
0.2
0.4
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Chromosome


C)
F_{ST}
0.0
0.2
0.4
0.6
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Chromosome


D)
Cluster A
471 regions
302 genes
789 regions
295 genes
188
5
257
4
105
29
461
Cluster B
Cluster C
881 regions
599 genes

E)
tripeptide transmembrane transporter activity
transmembrane transport
transmembrane receptor protein serine/threonine kinase binding
transferase activity, transferring glycosyl groups
stamen filament development
solute:proton antiporter activity
signaling receptor activity
RNA helicase activity
response to wounding
response to salicylic acid
response to jasmonic acid
quercetin 7−O−glucosyltransferase activity
quercetin 3−O−glucosyltransferase activity
proton transmembrane transport
proteolysis involved in cellular protein catabolic process
protein targeting to peroxisome
protein phosphorylation
protein import into peroxisome matrix
protein histidine kinase activity
protein binding
pollen tube
plasma membrane
phosphorelay sensor kinase activity
nucleolus
mRNA binding
membrane
mannan endo−1,4−beta−mannosidase activity
manganese ion binding
lysosome
lipid biosynthetic process
lignin biosynthetic process
ligand−gated ion channel activity
intracellular ligand−gated ion channel activity
hyperosmotic salinity response
hormone−mediated signaling pathway
glutathione binding
glutamate receptor activity
glucuronosyltransferase activity
gibberellin catabolic process
extracellular space
embryo sac development
cytosol
cytoplasm
cysteine−type endopeptidase activity
chloroplast stroma
chloroplast
cellular response to amino acid stimulus
cellular calcium ion homeostasis
auxin polar transport
auxin homeostasis
ADP binding
GO Terms
0.00 0.25 0.50 0.75 1.00
Enrich ratio
p−value
8 x 10⁻⁴
6 x 10⁻⁴
4 x 10⁻⁴
2 x 10⁻⁴
0
Gene number
30
60
90


A)
Orthology analysis
Pentose and glucuronate interconversions
150 genes (73 EC)
287 genes (73 EC)
57 Orthogroups
Starch and sucrose metabolism
136 genes (76 EC)
148 genes (76 EC)
76 Orthogroups
Flavonoid biosynthesis
83 genes (28 EC)
129 genes (28 EC)
28 Orthogroups
Additional genes found by orthology analysis
D)
Fst (1,053)
Pathways (564)
1,026
27
537
B)
Intersection Size
Musa acuminata
Dioscorea dumetorum
Elaeis guinensis
Dioscoread alata
Dioscorea zingiberensis
Trichopus zeynalicus
Phalaenopsis equestris
Colocasia esculenta
Phoenix dactylifera
Dioscorea rotundata
Oropetium thomaeum
Brachypodium distachyon
Sorghum bicolor
Setaria italica
Zea mays
Hordeum vulgare
Oryza sativa
Triticum aestivum
C)
Intersection Size
Dioscorea dumetorum
Dioscorea alata
Dioscorea zingiberensis
Trichopus zeylanicus
Dioscorea rotundata


A
S19_17029453 (G/C)
--CAATAGTAG
--CAATACTAG
LFKTQ-
LFKTQY-
5'-UTR
Exon
3'-UTR
Flavanone 3-hydroxylase
Dioal.19G083800
S4_16031453 (C/T)
--TTTGGCTT--
--TTTAGCTT--
--GCAVW--
--GCAV-
UDP-glucuronosyltransferase
Dioal.04G073400
S4_20715506 (C/T)
---TLH
---TLW
Pectin methylesterase
Dioal.04G130600
B
S4_20715506
P=0.008*
Gumminess
C/C
C/T
C
S4_20715506
P=0.008*
Total area (N.sec)
C/C
C/A
D
S19_17029453 (G/C)
P=0.035*
Brown index
G/G
C/G
E
Dioal.19G083800.1
Expression (FPKM)
Roujol62
Roujol75
F
S4_16031453 (C/T)
P=0.022*
Hue index
C/C
C/T

Trait	SNP_ID	Chr	Pos (bp)	Allele		Consequence Type	Genomic Location	Effect (%)	Gene Model	Predicted Annotation	Pathway	QTN
				Major	Minor							
Springiness	S1_10223458	1	10223458	C	T	Stop gain	p.Trp287*	0.38	*Dioal.01G024400.1*	Leucine-rich repeat	Pectin	*qSp1.1*
	S1_10224508	1	10224508	G	A	Stop gain	p.Gln104*	0.36	*Dioal.01G024400.1*	Leucine-rich repeat	Pectin	*qSp1.2*
	S2_19082482	2	19082482	C	T	Stop gain	p.Gln1032*	-0.05	*Dioal.02G062700.1*	Protein kinase	Pectin	*qSp2.1*
	S2_19095245	2	19095245	C	G	Stop gain	p.Tyr578*	-0.04	*Dioal.02G062800.1*	Protein kinase	Pectin	*qSp2.2*
	S13_11781422	13	11781422	A	G	splice_acceptor_variant&intron_variant	c.63-2A>G	-0.37	*Dioal.13G053400.1*	Xylose isomerase	Pectin	*qSp13.1*
	S20_14416887	20	14416887	G	T	Stop gain	p.Gly381*	-0.46	*Dioal.20G055400.1*	Malectin-like	Pectin	*qS20.1*
Gumminess	S4_20715506	4	20715506	C	T	Start Loss	p.Met1?	-3.97	*Dioal.04G130600.1*	Pectin methylesterase	Pectin	*qG4.1*
	S7_6378890	7	6378890	G	A	splice_donor_variant&intron_variant	c.-109+1G>A	-9.23	*Dioal.07G051400.1*	Alpha amylase	Starch	*qG7.1*
	S13_11783621	13	11783621	G	T	Stop gain	p.Glu208*	-2.61	*Dioal.13G053400.1*	Xylose isomerase	Pectin	*qG13.1*
	S13_11784207	13	11784207	A	G	splice_acceptor_variant&intron_variant	c.723-1G>A	23.60	*Dioal.13G053400.1*	Xylose isomerase	Pectin	*qG13.2*
	S19_24800556	19	24800556	G	A	splice_donor_variant&intron_variant	c.537+2C>T	4.87	*Dioal.19G187400.1*	Pectinesterase	Pectin	*qG19.2*
Total area	S2_18814720	2	18814720	T	A	Stop gain	p.Tyr289*	21.88	*Dioal.02G061200.1*	Protein kinase	Pectin	*qTa2.1*
	S4_20715506	4	20715506	C	T	Start Loss	p.Met1?	-27.90	*Dioal.04G130600.1*	Pectin methylesterase	Pectin	*qTa4.1*
	S7_16450864	7	16450864	G	T	Stop gain	p.Tyr228*	-29.42	*Dioal.07G067100.1*	4-Alpha-glucanotransferase	Starch	*qTa7.1*
Hardness	S13_28669451	13	28669451	C	G	splice_donor_variant&intron_variant	c.-837+1G>C	62.22	*Dioal.13G093500.1*	Beta-glucosidase	Starch	*qH13.2*
	S20_14334067	20	14334067	G	T	Stop gain	p.Gly6*	-12.80	*Dioal.20G055000.1*	Malectin-like	Pectin	*qH20.1*
	S20_14364743	20	14364743	A	C	Stop loss	p.Ter592Serext*?	10.75	*Dioal.20G055100.1*	Malectin-like	Pectin	*qH20.2*
Starch Content	S10_2314181	10	2314181	G	T	Stop gain	p.Gly35*	-14.75	*Dioal.10G014800.1*	1,3-beta-Glucan synthase	Starch	*qSC10.1*
	S15_23348104	15	23348104	T	C	splice_donor_variant&intron_variant	c.1029+2C>T	4.99	*Dioal.15G119500.1*	beta-Amylase	Starch	*qSC15.2*
	S17_19982533	17	19982533	T	C	splice_acceptor_variant&intron_variant	c.811-2A>G	-9.70	*Dioal.17G112400.1*	Glucose-1-phosphate adenylyltransferase	Starch	*qSC17.1*
Brown Index	S19_17029453	19	17029453	G	C	Stop gain	p.Tyr339*	34.69	*Dioal.19G083800.1*	Flavanone 3-dioxygenase	Flavonoid	*qYI19.1*
HUE Index	S4_16031453	4	16031453	C	T	Stop gain	p.Trp272*	50.28	*Dioal.04G073400.1*	UDP-glucosyltransferase protein	Flavonoid	*qHI4.1*