bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

1 **Genome-wide association analysis reveals new insights into the genetic architecture of**
2 **defensive, agro-morphological and quality-related traits in cassava**
3

4 Ismail Yusuf Rabbi[1*], Siraj Ismail Kayondo[1], Guillaume Bauchet[2], Muyideen Yusuf[1], Cynthia
5 Idhigu Aghogho[1], Kayode Ogunpaimo[1], Ruth Uwugiaren[1], Ikpan Andrew Smith[1], Prasad Peteti[1],
6 Afolabi Agbona[1], Elizabeth Parkes[1], Ezenwaka Lydia[3], Marnin Wolfe[5], Jean-Luc Jannink[5,6],
7 Chiedozie Egesi[1,3,4], and Peter Kulakow[1]
8 Author affiliations
9 [1] International Institute of Tropical Agriculture (IITA), Ibadan, Oyo State, 200001 Nigeria,
10 [2] Boyce Thompson Institute, Ithaca, New York 14853;
11 [3] National Root Crops Research Institute (NRCRI), PMB 7006, Umudike, 440221 Nigeria;
12 [4] Global Development Department, College of Agriculture and Life Sciences, Cornell Univ.,
13 Ithaca, New York 14850;
14 [5] Section on Plant Breeding and Genetics, School of Integrative Plant Sciences, Cornell Univ.,
15 Ithaca, NY 14850;
16 [6] United States Department of Agriculture - Agriculture Research Service, Ithaca, New York
17 14850.
18
19 * Corresponding author: I.Rabbi@cgiar.org; ORCID ID: 0000-0001-9966-2941
20

21 **Abstract**
22 Cassava (*Manihot esculenta*) is one of the most important starchy root crops in the tropics due to
23 its adaptation to marginal environments. Genetic progress in this clonally propagated crop can be
24 accelerated through the discovery of markers and candidate genes that could be used in cassava
25 breeding programs. We carried out a genome-wide association study (GWAS) using a panel of
26 5,310 clones developed at the International Institute of Tropical Agriculture - Nigeria. The
27 population was genotyped at more than 100,000 SNP markers via genotyping-by-sequencing
28 (GBS). Genomic regions underlying genetic variation for 14 traits classified broadly into four
29 categories: biotic stress (cassava mosaic disease and cassava green mite severity); quality (dry
30 matter content and carotenoid content) and plant agronomy (harvest index and plant type). We also
31 included several agro-morphological traits related to leaves, stems and roots with high heritability.
32 In total, 41 significant associations were uncovered. While some of the identified loci matched
33 with those previously reported, we present additional association signals for the traits. We provide
34 a catalogue of favourable alleles at the most significant SNP for each trait-locus combination and

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

35 candidate genes occurring within the GWAS hits. These resources provide a foundation for the
36 development of markers that could be used in cassava breeding programs and candidate genes for
37 functional validation.

38

39 **Keywords:**

40 Cassava, breeding, genome-wide association, pest and disease resistance, agro-morphology and
41 root quality traits, genotyping-by-sequencing

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

42 **1.0 Introduction**

43 Cassava (*Manihot esculenta* Crantz) is not only the most widely consumed starchy-root staple but
44 also an emerging multi-purpose and industrial crop in Africa, Asia, and Latin America (Parmar et
45 al. 2017). This clonally propagated species show remarkable adaptation to diverse agro-ecologies
46 and can produce reasonable yield under marginal conditions of climate and soil (Jarvis et al. 2012).
47 In addition, its flexible harvest window allows the crop to be left in the soil as a food reserve.
48 These properties make cassava an ideal food security crop with an increasing trend in global
49 production (Prudencio and Al-Hassan 1994; Burns et al. 2010).
50
51 In the last four decades, cassava breeding programs across Africa, Asia, and Latin America, have
52 developed varieties that can withstand production constraints including biotic and abiotic stresses
53 with improved yield and starch content (Kawano 2003; Okechukwu and Dixon 2008). While
54 phenotype-based recurrent selection has made significant progress, the rate of genetic gain has
55 been low due to several breeding complexities associated with the biology of the crop, including
56 asynchronous flowering, low seed set per cross, a long cropping cycle of 12 to 24 months and low
57 multiplication rate of planting materials (Ceballos et al. 2012). These challenges hinder the
58 breeding program's abilities to rapidly respond to changing human needs under volatile climatic
59 and environmental conditions.
60
61 Modern breeding methods including marker-assisted selection (MAS) and genomic selection (GS)
62 can be used to accelerate genetic improvement particularly by reducing generational interval and
63 increasing selection intensity (Ferguson et al. 2012; Ceballos et al. 2015; García-Ruiz et al. 2016).
64 However, integration of molecular markers as part of MAS in breeding pipelines requires an initial
65 investment in discovery research to identify major-effect loci that serve as the targets of selection.
66 With the rapid advances in next-generation sequencing (NGS) technologies, it is now feasible to
67 generate genome-wide marker data in large populations. This, coupled with phenotype data makes
68 it possible to identify and map locations of agriculturally important genes and quantitative trait
69 loci (QTL) at the whole genome level (Varshney et al. 2014).
70
71 Significant investment has been made in the development of genomic resources for cassava
72 including dense multi-parental linkage maps (International Cassava Genetic Map Consortium
73 (ICGMC), 2015), annotated reference genomes (Prochnik et al. 2012; Zhang et al. 2018) and a
74 haplotype map of common genetic variants from deep sequencing of hundreds of diverse clones
75 (Ramu et al. 2017). Several genome-wide association studies (GWAS) have been conducted to

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

76 describe the genetic architecture of resistance against cassava mosaic disease (Wolfe et al. 2016),
77 reduced green mite infestation (Ezenwaka et al. 2018), cassava brown streak disease (Kayondo et
78 al. 2018) and provitamin A and dry matter content (Esuma et al. 2016; Rabbi et al. 2017; Ikeogu
79 et al. 2019).
80
81 Using a collection of 5,130 elite cassava clones derived from three cycles of recurrent selection in
82 the International Institute of Tropical Agriculture (IITA) Cassava Breeding Program, we examined
83 the genetic architecture of 14 continuous and categorical traits, including defense against biotic
84 stresses, agro-morphological and quality-related traits (Table 1). Among the biotic stress-related
85 traits, we considered cassava mosaic disease (CMD) and cassava green mite (CGM). Caused by
86 different species of cassava mosaic geminiviruses, CMD is one of the most important biotic
87 constraints to cassava production in Africa, India and Sri Lanka (Thottappilly et al. 2003; Alabi et
88 al. 2015; CABI 2019) and has recently spread to the South Asian countries of Thailand and
89 Vietnam (Uke et al. 2018). Infected plants can incur yield losses of up to 82% which translates to
90 more than 30 million tonnes of fresh cassava roots loss annually (Owor et al. 2004; Legg et al.
91 2006). Infestation by CGM (*Mononychellus tanajoa*) during the dry season causes chlorosis and
92 restricted growth resulting in a significant negative impact on root yield. The main type of
93 resistance to CGM is attributed to apical leaf pubescence but may also include other mechanisms
94 (Shukla 1976; Byrne et al. 1982; Ezenwaka et al. 2018).
95
96 For quality traits, we considered total carotenoid content and dry matter content variation.
97 Biofortification breeding to increase provitamin A carotenoids in storage roots is an important goal
98 in cassava improvement programs around the world (Saltzman et al. 2013). Although the crop's
99 gene pool contains accessions that are naturally rich in provitamin A, they make up a small
100 proportion of cultivated varieties especially in Africa (Welsch et al. 2010). We used the colour-
101 chart based assessment of total carotenoid content variation. In cassava, the intensity of root
102 yellowness is strongly correlated with total carotenoid content (Sánchez et al. 2014). Additionally,
103 provitamin A is the predominant carotenoid component in cassava (Ceballos et al. 2017). Dry
104 matter content is a key yield component that determines varietal acceptance by farmers and
105 processors (Bechoff et al. 2018). Cassava germplasm contains considerable variation in percentage
106 dry matter content of the fresh roots ranging from as low as 10% to 45% (Kawano et al. 1987).
107
108 Among agronomic traits, harvest index and plant type were included in the present study. Harvest
109 index reflects the partitioning of resources between the storage roots and above-ground biomass

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

110 (Sinclair 1998). The desirable harvest index values for the crop range from 0.5 to 0.7 (Kawano et
111 al. 1978; Kawano et al. 1998). Plant type in cassava can be characterized by four general
112 descriptive shapes, namely cylindrical, umbrella, compact, and open. Plants with cylindrical
113 shapes do not form branches and are most preferred for mechanized farming. Umbrella plant types
114 generally branch at a high level (above 1 m) and have fewer levels of branching whereas compact
115 and open types are characterized by low first branching height and multiple branching levels but
116 differ in the angle of branches and erectness of the stems (Fukuda et al. 2010).
117
118 We also assessed the genetic architecture of several morphological traits related to leaves (petiole
119 colour, apical leaf colour, mature leaf greenness, leaf shape) and storage roots (periderm and cortex
120 colour). Colour of cassava leaf petioles as well as apical leaves ranges from light green to purple
121 due to anthocyanin pigmentation. Anthocyanins play various roles in plants including protection
122 against ultraviolet light, overcoming different abiotic and biotic stresses and other physiological
123 processes such as leaf senescence (Gould et al. 2008). Mature leaf greenness is related to
124 chlorophyll content, an indicator of a plant's photosynthetic capacity (Palta 1990). Additionally,
125 cassava germplasm exhibits diverse leaf shapes ranging from ovoid lobes to linear forms (Fukuda
126 et al. 2010). This variation is useful as a morphological descriptor and could also play other
127 functional role related to light capture (Takenaka 1994). Cassava root periderm colour varies from
128 cream through light brown to dark brown while that of the cortex includes cream, pink and purple.
129 The few but predominant farmer-preferred varieties in Africa often have a pink or purple cortex
130 and dark brown periderm although there is no proven genetic correlation between these traits and
131 the culinary properties in cassava. On the other hand, industrial processing into starch and flour
132 production generally utilize whole roots after mechanical periderm removal. These industries
133 prefer varieties with white/cream periderm in order to ensure bright-coloured products. Finally,
134 we also assessed the genetic architecture of stem colour, another morphological descriptor used
135 for variety identification and varies from orange to dark-brown (Fukuda et al. 2010).
136
137 Here, we provide a catalogue of genomic loci associated with the 14 traits, a list of favourable
138 alleles for each trait-locus combination and candidate genes located within the identified loci.
139 Understanding the genetic architecture of variation in the studied traits is an important step towards
140 the development of molecular tools to accelerate the transfer of favourable alleles into farmer-
141 preferred varieties.
142

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

143 **Material and Methods**

144 **Field experiments and phenotyping**

145 The breeding population composed of 5,130 elite IITA cassava breeding genotypes were
146 phenotyped for the 14 traits across four contrasting locations in Nigeria; Ubiaja (6°40′ N, 6°20' E),
147 Ibadan (7°24′ N, 3°54′ E), Mokwa (9°21' N 5°00' E) and Ikenne (6°52' N 3°42'E) from 2013 to
148 2016. Description of the traits, their ontologies and measurement methods are available at
149 https://cassavabase.org/search/traits. This population consisted of 717 elite lines from the genetic
150 gain (GG) population, 2,322 full-sib progenies derived from 88 elite GG progenitors (Cycle1 –
151 C1), and 2,091 full-sib progenies derived from top 89 cycle one progenitors (Cycle2 – C2)
152 (Supplementary Table S1). The mean family size for C1 clones was 15, ranging from 1 to 77 clones
153 while that of C2 was 7.6 ranging from 1 to 20 clones. The GG population is a collection of the
154 "Tropical *Manihot* Selection (TMS)" cultivars developed in the past five decades by IITA (Dixon
155 and Ssemakula 2008; Rabbi et al. 2017). This panel represents an extensive and diverse pedigree
156 comprising of crosses between West African landraces, Latin American elite and wild germplasm.
157 The raw phenotypic data is openly accessible on CassavaBase repository
158 (ftp://ftp.cassavabase.org/manuscripts/PlantMolBiol_Rabbi_et_al_2020/)

159

160 The GG population was evaluated in clonal evaluation trials (CET) while the C1 and C2 were
161 evaluated in CET, preliminary yield trials (PYT), and advanced yield trials (AYT). The CET plots
162 were composed of a single row with 5 plants per plot, in an augmented block design, with 2 checks
163 as a control. PYT plots consisted of two rows with 10 plants per plot, in a randomized complete
164 block design with 2 reps and 5 checks as a control. AYT plots contained four rows with 20 plants
165 per plot, in a randomized complete block design with 3 reps and 5 checks as a control. Planting
166 was performed from June to July (during the rainy season) and harvested around the same time the
167 following year. Spacings between rows and plants were 1 and 0.8 m in all trials, respectively,
168 except in CETs where we used 0.5 m within rows. All field trial management was performed,
169 whenever necessary, in accordance with the technical recommendations and standard agricultural
170 practices for cassava (Fukuda et al. 2010; Abass et al. 2014; Atser et al. 2017).

171

172 **Genotyping**

173 Genomic DNA was extracted from freeze-dried leaf samples following a modified Dellaporta
174 CTAB method (Dellaporta et al. 1983). DNA quality and quantity were assessed using a Nanodrop
175 1000 spectrophotometer at 260 nm absorbance. Genome-wide single nucleotide polymorphism
176 (SNP) data was generated using the genotyping-by-sequencing (GBS) approach described by

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Elshire et al. (2011). Reduction in genome complexity for GBS was achieved through restriction digestion using ApeKI enzyme (Hamblin and Rabbi 2014). Sequencing reads were aligned to the cassava V6 reference genome (Prochnik et al. 2012) followed by SNP calling using TASSEL GBS pipeline V4 (Glaubitz et al. 2014). SNP calls with less than 5 reads were masked before imputation using Beagle V4.1 (Browning 2016). A total of 202,789 biallelic SNP markers with an estimated allelic r-squared value (AR2) of more than 0.3 were retained for subsequent analyses after imputation.

Statistical analyses

Phenotype data analyses

Our interest in this study was to identify the genetic architecture of selected traits rather than location- or year-specific QTLs. We collapsed plot observations for each genotype to a single best linear unbiased prediction (BLUP) using the following mixed linear model (MLM) with the lme4 (Bates et al. 2015) package in R (R Development Core Team 2016): $y_{ij} = \mu + g_i + \beta_j + r_{j(l)} + \varepsilon_{ijl}$ where y_{ij} represents vector of phenotype data, μ is the grand mean, g_i is the random effect of genotype i with $g_i \sim N(0, \sigma_g^2)$, β_j are the fixed effects of year – location combination j, $r_{j(l)}$ is a random effect of replication nested within location-year combination assumed to be distributed $N(0, \sigma_r^2)$; and ε_{ijl} is the residual with $\varepsilon_{ijl} \sim N(0, \sigma_e^2)$.

Pairwise correlation between traits was determined from BLUPs using the R function "*cor*" in the "*stats*" package (R Development Core Team 2016) and visualization of the correlation matrices was done using the '*corrplot*' R package (Wei and Simko 2017). Due to the unbalanced nature of trials, we calculated the broad-sense heritability estimates on a plot-mean basis using the formula $H^2 \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$. Additive genomic heritability for each trait was also estimated using a linear mixed model as implemented in GCTA (Yang et al. 2011; Speed et al. 2012).

Population structure and genetic relatedness assessment

Population structure and cryptic genetic relatedness are known to confound GWAS analysis and lead to spurious associations (Chen et al. 2016). To assess the extent of population structure, we performed a principal component analysis (PCA) using PLINK-v1.9 (Purcell et al. 2007; Rentería et al. 2013). The pairwise Kinship matrix was visualized using the heatmap function in R.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

210 **Genetic architecture of studied traits**

211 For each trait, single-locus GWAS analysis was carried out using the mixed linear model (MLM)
212 approach implemented in Genome-Wide Complex Trait Analysis (GCTA) software (Yang et al.
213 2011). Because the inclusion of the candidate marker in kinship calculation when controlling for
214 cryptic relatedness can lead to a loss in power (Yang et al. 2011; Listgarten et al. 2012), we also
215 used the mixed linear model which excluded candidate markers via a leave-one-chromosome-out
216 analysis as implemented in GCTA (Yang et al. 2011; Yang et al. 2014). The first approach is
217 referred to as MLMi ("i" for candidate marker included) and the second approach is referred to as
218 MLMe ("e" for candidate marker excluded). Visualisation of MLMi and MLMe results in the form
219 of Manhattan and quantile-quantile plots were implemented in the '*CMplot*' R package (LiLin-Yin
220 2019).

221

222 **SNP Marker and favourable allelic prediction**

223 To assess the genetic architecture of the studied traits, we fit a linear regression model using peak
224 SNPs at the identified loci as independent variables against the traits' BLUPS as the response
225 variables. The relative allele substitution effects at each marker were visualized using boxplots.
226 Here, a locus was defined as a uniquely identifiable genomic region whose SNPs passed genome-
227 wide Bonferroni significance threshold (α=0.05/101,521 = 4.93 × 10-7).

228

229 **Candidate gene identification**

230 Candidate loci were explored using a combination of GWAS p-values, local linkage disequilibrium
231 (measured as r^2), and gene annotation using the gff3 file of the cassava genome available on
232 phytozome v.12.1 (https://phytozome.jgi.doe.gov/pz/portal.html) (Goodstein et al. 2011; Batra et
233 al. 2014). This information set was summarized for each candidate loci using Locus zoom (Pruim
234 et al. 2010). To obtain a regional zoom plot for each candidate locus, we built a local SQLite
235 database including the 101,521 biallelic SNP marker matrix, associated GWAS p-values for each
236 trait analyzed and the cassava gene-annotation following instructions available at
237 https://genome.sph.umich.edu/wiki/LocusZoom_Standalone. Gene codes were shortened to ease
238 visualization and whenever available, *Arabidopsis thaliana* homologues were noted.
239 Recombination information was provided using the same approach as described in Wolfe et al.
240 (2016). A standard interval of 100 Kb (50 kb upstream, 50 kb downstream) was explored for each
241 candidate locus and adjusted according to the extent of local linkage disequilibrium with the
242 candidate SNP (r^2>0.8).

243

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

244 **Results**

245 **Variation and relationships among traits**

246 The frequency distributions of the BLUPs per trait within the panel are presented in Supplementary
247 Figure S1 and the raw BLUPs values are also presented in Supplementary Table S1. An analysis
248 of the phenotypic classes of the panel showed that almost all measured traits followed a normal
249 distribution. A couple of traits were slightly skewed towards the tails. The heritability and variance
250 components estimates associated with the studied traits are presented in Table 2. All measured
251 traits exhibited larger than 2 fold differences between the maximum and minimum values with a
252 mean coefficient of variation (CV) of 35%. Apical pubescence was the only trait with a
253 significantly larger CV of 113%. These large differences between the maximum and minimum
254 values are an indication of broad genetic variability within the mapping panel. To estimate the
255 influence of additive and non-additive genetic effects on the observed phenotypic values, we
256 estimated broad-sense and genomic-heritabilities. Both the broad-sense and SNP-based heritability
257 estimates were comparable ranging from low to high: 15 to 78% and 17 to 72%, respectively.
258 Some traits like total carotenoids, dry matter, petiole colour, apical leaf colour, and stem colour
259 had higher SNP-heritability estimates (> 0.5) relative to other traits like harvest index, plant type,
260 and resistance to cassava green mite with the lower heritabilities (< 0.4). SNP-based heritability is
261 useful in approximating the proportion of phenotypic variance attributable to the additive genetic
262 variation (Yang et al. 2017). The moderate to high levels of SNP-based heritabilities found for
263 traits under active selection coupled with sufficient variability in the population indicates good
264 potential for genetic improvement of these traits through recurrent selection.

265

266 The relative magnitude of phenotypic correlation pairs ranged from 0.27 between total carotenoid
267 variation by colour chart (TC-chart) and root cortex colour to -0.65 for mature leaf greenness and
268 petiole colour (Figure 1). We detected highly significant negative phenotypic correlations between
269 dry matter and total carotenoid contents (-0.31, $P < 0.001$); mean CMD severity and harvest index
270 ($r = -0.14$, $P < 0.001$); and apical pubescence and CGM severity ($r = -0.27$, $P < 0.001$). There were
271 highly significant positive correlations between total carotenoids and periderm colour ($r = 0.27$,
272 $P<0.001$), leaf greenness ($r = 0.17$, $P < 0.001$), and stem colour ($r = 0.12$, $P < 0.001$).

273

274 **Distribution of SNP Markers and Population stratification**

275 Genotyping of the panel enabled the identification of 202,789 imputed SNP markers. Upon
276 filtering markers with minor allele frequency less than 1% we retained 101,521 SNPs. All the
277 101K SNPs were mapped onto the 18 chromosomes covering a total of 532.5 Mb of the cassava

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

278 genome and a SNP density of 5.6 variants/Kbp (Figure 2). Individually, the SNP coverage per
279 chromosome ranged from 3,821 on chromosome 7 to 11,189 on chromosome 1. The average minor
280 allele frequency was 0.196 and 80% of the SNPs had frequency greater than 5% indicating
281 enrichment of common SNP alleles in the population (Figure 2). To estimate the mapping
282 resolution for our panel, we assessed pairwise linkage disequilibrium (LD) between 101,521 SNPs
283 across the 5,130 cassava accessions. We used the mean r^2 value as an estimate of LD decay using
284 a window of 1 Mbp, followed by fitting a non-linear regression curve of LD versus distance. The
285 whole-genome LD decay peaked at r^2 of 0.349 and dropped to an r^2 of 0.212 at a distance of 10
286 Kb (Figure 2).
287
288 Principal component analysis (PCA) was conducted to visualize the extent and degree of
289 population structure present within the panel. While the first two principal components (PC)
290 explained 7.4% of the total phenotypic variation, PC3 to PC10 together explained about 14% of
291 the total variation (Figure 3). We observed a considerable overlap between cycles (C0, C1 and
292 C2); hence, no distinct clusters were detected in the panel. The results of the identity-by-state
293 distance showed a considerable range of relationship in the association panel with an average value
294 of 0.23 and ranged from 0.02 to 0.32. We observed familial relationships along the diagonal with
295 a few large blocks of closely related individuals. The off-diagonal indicated low kinship (Figure
296 3).
297

298 **Genome-wide association results**

299 We identified a total of 27 unique genomic regions significantly associated with variation in the
300 14 studied traits following the MLMi analysis (Figure 4). Additional loci were uncovered for a
301 majority of the traits when we considered the MLMe approach, bringing the total number of loci
302 to 41 (Supplementary Figure S2). The most significant trait-marker associations from MLMi and
303 MLMe for each trait and genomic region combinations are also provided in Table 3. In the
304 following sections, we present the results and provide discussion for each trait starting with
305 economically important traits to morphological traits.
306

307 **Cassava mosaic disease severity**

308 The genetic basis of CMD resistance has been studied extensively using bi-parental linkage
309 mapping (Rabbi et al. 2014) and GWAS (Wolfe et al. 2016). These studies repeatedly showed that
310 the main resistance to the disease is conferred by a single major gene on chromosome 12 which is
311 widely known as CMD2 locus (Akano et al. 2002). In the present study, we uncovered the same

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

312 locus and two other loci on chromosome 14. The CMD2 region on chromosome 12 is tagged by
313 marker S12_7926132 (p-value = 1.7 $\times 10^{-112}$). The favourable allele at this marker (T) was found
314 to occur at high frequency in the population (Freq > 0.66). The two additional loci on chromosome
315 14 are tagged by markers S14_4626854 (p-value = 1.7 $\times$ 10^{-14}) and S14_9004550 (p-value = 4.2 $\times$
316 10^{-17}). The SNP at the CMD2 locus had a larger effect (β value 0.82) compared to the two new
317 loci on chromosome 14 (β value of 0.23 and 0.28).
318
319 **Cassava green mite severity**
320 We found four genomic regions associated with CGM resistance in our panel (Figure 4) of which
321 only marker S8_6409580, occurring around 6.41 Mb region of chromosome 8, was previously
322 reported (Ezenwaka et al. 2018). The remaining loci were found on chromosomes 1, 12 and 17
323 (Table 3). We note that except for the markers on chromosome 8, which were significant in both
324 MLMi and MLMe, the remaining loci were only significant in the MLMe analysis, though their
325 SNP effects (β statistic) were very similar in the two analyses. Association analysis of the related
326 trait, apical pubescence identified significant loci on five chromosomes, two of which co-locate
327 with same regions underlying resistance to CGM on chromosomes 8 (S8_6409580) and 12
328 (S12_5524524). The genetic correlation between the two traits is -0.51 in the population
329 (Supplementary Table S2). The other loci were on chromosomes 9 (S9_1588034), 11
330 (S11_5727254), and 16 (S16_1501762).
331
332 **Carotenoid content**
333 GWAS analysis of colour-chart based variation in root carotenoid content revealed a major locus
334 on chromosome 1 tagged by three markers around 24.1, 24.6 and 30.5 Mb regions with the top
335 markers being S1_24159583, S1_24636113 and S1_30543962, respectively. Previous GWAS
336 analyses reported significant associations in this region (Esuma et al. 2016; Rabbi et al. 2017;
337 Ikeogu et al. 2019). In addition, we uncovered five new genomic regions associated with this trait
338 on chromosome 5 (around 3.38 Mb), 8 (two peaks at 4.31 and 25.59 Mb regions), and 15 (7.65
339 Mb), and 16 (0.48 Mb). We note that the regions in the last three chromosomes, including two on
340 chromosome 8, were only detected via the MLMe analysis.
341
342 **Dry matter content**
343 GWAS for variation in dry matter content following the MLMi analysis revealed two major loci
344 for DM of which one was previously reported (Rabbi et al. 2017). The most significant locus

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

345 occurs on chromosome 1 around 24.64 Mb region and is tagged by marker S1_24636113 (p-value
346 = 5.0×10^{-8}). The second locus was tagged by marker S6_20589894 (p-value = 3.4×10^{-7}).
347 Additional loci on chromosomes 12 (S12_5524524, p-value = 8.0×10^{-12}), 15 (S15_1012346, p-
348 value = 4.0×10^{-17}), and 16 (S16_25663808, p-value = 4.2×10^{-12}) were detected from the MLMe
349 analysis. Along with the new locus on chromosome 6, these regions have not been previously
350 reported to be associated with dry matter content variation and are suitable for further genetic
351 studies including identification of underlying candidate genes.
352

353 **Harvest index**

354 The MLMi-based GWAS for harvest index, the ratio of fresh root weight to total plant weight,
355 uncovered two genomic regions that are significantly associated with the trait. The first peak is in
356 chromosome 2, tagged by SNP S2_2809137 (p-value of p-value = 3×10^{-8}). The second locus
357 occurred on chromosome 12 with SNP S12_6055806 showing the strongest association with the
358 trait (p-value = 5.4×10^{-24}). Analysis of the same trait using the MLMe approach uncovered several
359 other regions scattered across chromosomes 3, 4, 6, 8, 9, 14, 15, 16, and 18.
360

361 **Morphological traits**

362 GWAS for outer cortex colour uncovered two association signals located on chromosomes 1 (3.05
363 Mbp region) and 2 (6.56 Mbp region) and tagged by SNPs S1_3047840 (p-value 1.6×10^{-92}) and
364 S2_6566608 (p-value 5.0×10^{-83}), respectively. A single genomic region on chromosome 3 (4.54
365 Mbp) was found to be linked to periderm colour. The most significant SNP at this locus
366 (S3_4545411) had a p-value of 1.4×10^{-123}.
367

368 Two association signals were detected for plant type from MLMe while no marker passed the
369 Bonferroni threshold in the MLMi analysis. The detected loci jointly occurred on chromosome 1
370 at 2.19 Mbp region (tagged by SNP S1_3192405, p-value 3.82×10^{-8}) and 25.30 Mbp region
371 (S1_25303195, 5.25×10^{-14}). Three loci were found to be associated with stem colour variation,
372 one in chromosome 2 and two on chromosome 8. The most significant locus was around 13.6 Mbp
373 region of chromosome 8 and is tagged by marker S8_13604799 (p-value 8.3×10^{-69}). The other
374 two loci were tagged by SNPs S2_13928566 (p-value 3.6×10^{-14}) and S8_22630799 (p-value 3.3
375 $\times 10^{-17}$).
376

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

377 Genetic architecture of leaf morphology traits showed that they are controlled by one to three
378 major loci, indicating simple genetic architecture. We found a single genomic region around 23.45
379 Mbp of chromosome 1 to be associated with this trait and is tagged by SNP S1_23452638 (p-value
380 9.8×10^{-180}). Notably, the same exact SNP was found to underlie the variation in mature leaf
381 greenness. It is therefore not surprising that these traits are negatively correlated in our population.
382 SNP effect analysis showed that while allele “T” at S1_23452638 had a positive effect on petiole
383 colour, the same allele showed a negative effect on leaf greenness. Regression of the marker on
384 the traits for leaf colour and petiole colour returned an R^2 0.57 and 0.62, respectively.
385
386 Variation in the colour apical leaves was found to be associated with 3 loci occurring on
387 chromosomes 2, 3, and 8. The most significant marker was S2_6086714 (p-value 6.1×10^{-83})
388 followed by the markers on chromosome 8 (S8_6061421, p-value 1.9×10^{-11}) and 3 (S3_4745233,
389 p-value 4.4×10^{-9}). Multiple regression returned an R^2 of 0.31 for this trait suggesting either a more
390 complex architecture or imprecise scoring of the variation in the trait. The GWAS result for leaf
391 shape uncovered two major loci on chromosome 15 occurring around 10.27 Mbp and 20.57 Mbp
392 regions. The first peak tagged by SNP S15_10273255 was highly significant (3.7×10^{-174}) while
393 the second peak was tagged by S15_20573383 (p-value 1.8×10^{-11}). Fitting linear model with the
394 two top SNPs for this trait returned R^2 of 0.40.
395

396 **Genetic architecture of studied traits**

397 To assess the genetic architecture of the studied traits, we fit a linear model using peak SNPs at
398 the identified loci (Table 3) as independent variables against the traits BLUPS as the response
399 variables. Peak SNPs at the identified loci explained approximately 34% of the trait variation on
400 average and R^2 ranged from 5% to 62% (Figure 5). Markers for cassava green mite severity, harvest
401 index, plant type and dry matter content had the lowest predictive ability ($R^2 < 0.1$). Most of the
402 morphology and colour related traits for leaves and roots had between 1 and 3 peaks of association
403 except apical leaf pubescence with 6 loci. Peaks associated with variation in total carotenoid
404 content and resistance to CMD had a large effect ($R^2 = 0.60$ and 0.45, respectively). The major
405 loci controlling these traits had known candidate genes reported previously. Still, the new loci
406 identified in these as well as the other traits are attractive candidates for follow-up studies.

407 **Identification of favorable alleles**

408 To identify favourable alleles for traits under selection in the breeding pipeline, the most
409 significantly associated SNP (lowest p-value) at each major-effect locus were chosen. Allelic

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

410 substitution effects for these markers are shown in Supplementary Figure S3. Selected traits
411 include resistance to CMD and CGM, increased dry matter and carotenoid content. For CMD
412 resistance, the haplotypes at the top SNPs S12_7926132 (allele G/T), S14_4626854 (A/G) and
413 S14_9004550 (T/C) are *T-A-C*, respectively. There were 251 genotypes that were fixed for the
414 favourable alleles in the population. Their average CMD severity BLUP was the lowest among all
415 haplotype combinations (mean = -0.56, SD = 0.29). We also note dominance allele effect at the
416 first SNP which is linked to the CMD2 locus, which agrees with previous results from biparental
417 QTL studies (Akano et al. 2002; Rabbi et al. 2014).
418
419 The CGM resistance-linked haplotype for SNPs S1_28672656 (A/T), S8_6409580 (C/G),
420 S12_5524524 (C/T) and S17_23749968 (A/G) are *T-C-C-G*, respectively. Alleles associated with
421 increased pubescence particularly for loci that co-locates with CGM resistance of chromosomes 8
422 (S8_6409580 (C/G)) and 12 (S12_5524524 (C/T)) are *C-T*.
423
424 Several SNP loci from chromosome 1 (S1_24159583, S1_24636113, S1_30543962), 5
425 (S5_3387558), 8 (S8_4319215, S8_25598183), 15 (S15_7659426), and 16 (S16_484011) were
426 associated with increased carotenoid content. The favourable haplotype for chromosome 1 SNPs
427 are *T-G-G*, respectively while that for chromosome 5 is *T*. The favourable haplotype for the two
428 loci on chromosome 8 is *A-T* while for chromosomes 15 and 16 SNPs it is *T-T*, respectively. No
429 individuals in the population were fixed for the favourable alleles across all the SNPs.
430
431 The favourable allele at the most significant dry matter locus on chromosome 1 (S1_24636113
432 allele G/A) is *A*. We note that this marker also co-located with the major locus for carotenoid
433 content and allele *A* has a non-favourable effect in carotenoid content. The favourable haplotypes
434 at the other loci on chromosomes 6 (S6_20589894 G/A), 12 (S12_5524524, C/T), 15
435 (S15_1012346 C/T) and 16 (S16_25663808 T/C) are *G-C-T-C*. A total of 61 genotypes fixed for
436 the favourable alleles at the SNPs on chromosomes 1, 6 and 12. Their average dry matter BLUP
437 was the highest among all haplotype combinations (mean = 2.62, SD = 3.66).
438
439 **Candidate gene identification**
440 The most significant GWAS peaks were further investigated for the presence of potential candidate
441 genes using local linkage disequilibrium, coupled to gene annotation. We matched the highly
442 significant SNP markers identified in the MLMe analysis with gene annotation in the regions up
443 and downstream derived from phytozome online database. Seventeen candidate genes that

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

444 colocalized with the identified putative genomic loci on height chromosomes were retrieved from
445 the cassava genome available on phytozome v.12.1 and are presented in Supplementary Table S3.
446 Several of these identified genes were further selected and highlighted based on their biological
447 significance within a given biological pathway. Additionally, we provide regional Manhattan plots
448 for each locus-trait combination in Supplementary Figure S4. These plots include candidate genes
449 within 100 Kb of the top SNP marker (50 kb upstream, 50 kb downstream) with some adjustments
450 based on the extent of local linkage disequilibrium with the candidate SNP.
451
452 Three genes associated with total carotenoid content (TC-chart) were identified on chromosomal
453 regions 1, 5 and 15. Of the three associated genes, Manes.01G124200 (Phytoene synthase)
454 occurred within the previously reported genomic region (Esuma et al. 2016; Rabbi et al. 2017).
455 Manes.01G124200 gene has a transferase enzymatic activity critical in the carotenoid biosynthesis
456 pathway. The other two genes are novel: Manes.05G051700 and Manes.15G102000 (both of
457 which are Beta-carotene dioxygenases) located at 3.87 Mb on chromosome 5 and 7.58 Mb on
458 chromosome 15, respectively, and are also known to play major roles in carotenoid biosynthesis.
459
460 For dry matter content, we identified two genes involved in starch and sucrose metabolism
461 occurring 600 Kbp away from top SNP S1_24636113 in chromosome 1: Manes.01G123000
462 (UDP-Glucose pyrophosphorylase) and Manes.01G123800 (Sucrose synthase) previously
463 reported (Rabbi et al. 2017). Although these genes are a few hundred Kb away from the top SNP,
464 this particular genomic region of chromosome 1 is known to harbour extensive LD in Africa
465 cassava germplasm (Rabbi et al. 2017). Other candidate genes found close to the top SNPs on
466 chromosomes 6, 15 and 16 are Manes.06G103600 (Bidirectional sugar transporter Sweet4-
467 Related); Manes.15G011300 (Sweet17 homologue, which mediates fructose transport across the
468 tonoplast of roots) and Manes.16G109200 (Hexokinase), respectively.
469
470 Our search for candidate genes related to CMD severity uncovered two peroxidase genes:
471 Manes.12G076200 and Manes.12G076300 occurring less than 45 Kbp away from marker
472 S12_7926132 in chromosome 12. These two candidate genes were previously reported by Wolfe
473 et. al. (2016) and Rabbi et al. (2014). Peroxidases have been reported to play a role in activating
474 plant defence systems upon pathogen infections (Ye et al. 1990; Wu et al. 1997; Gonçalves et al.
475 2013).
476

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

477 Our analyses further detected three loci that colocalized with three candidate genes associated with
478 CGM severity. Specifically, SNP marker S8_6409580 fell in the coding region of
479 Manes.08G058000 gene, a homolog of AtMYB16, encoding a MIXTA-like MYB gene which
480 regulates cuticle development and trichome branching (Oshima et al. 2013). Structural traits
481 including trichomes and waxy cuticles are known to act as a physical barrier to arthropod pest
482 attachment, feeding and oviposition (Mitchell et al. 2016)
483
484 Harvest index GWAS analysis highlighted two regions on chromosome 2 and 12, respectively. On
485 chromosome 2, the candidate SNP (S2_2809137) is located 24 Kbp away Manes.02G035900 a
486 homologue of BFRUCT4 (vacuolar invertase) a key enzyme in sucrose hydrolysis and involved in
487 the export of reduced carbon sink organs such as roots (Haouazine-Takvorian et al. 1997; Nägele
488 et al. 2010). For leaf shape, a candidate gene Manes.15G136200, homologous to KNOX1 that is
489 implicated in the expression of diverse leaf shapes in plants was found around 186 Kb away from
490 the major locus on chromosome 15 tagged by SNP S15_10273255 (Furumizu et al. 2015).
491 Candidate gene search around the major locus for petiole colour and leaf greenness tagged by SNP
492 S1_23452638 revealed the presence of a Myb transcription factor homologue Manes.01G115400
493 which occurred 30.7 Kb away from the top marker. Myb genes are known to play a key role in
494 regulating pigment biosynthesis pathway in plants (Nesi et al. 2001; Kobayashi et al. 2002; Himi
495 and Noda 2005; Allan et al. 2008; Furumizu et al. 2015)
496
497 **Discussion**
498 Understanding the genetic architecture of key breeding-goal traits is a critical step towards more
499 efficient and accelerated genetic improvement. This study builds on and extends previous cassava
500 GWAS efforts by analysing a large breeding population phenotyped extensively over successive
501 years and stages of selection in multi-environment field trials. The population showed large
502 phenotypic variation among clones with respect to all traits (Supplementary Figure S1).
503 Furthermore, the population was derived from two successive cycles of recurrent selection using
504 elite parents with good breeding values for yield, dry matter content and resistance against CMD
505 (Wolfe et al. 2016). The collection is therefore expected to be enriched for favourable alleles from
506 major- and minor-effect loci underlying these traits and is therefore well suited to efficiently
507 conduct a marker-trait association study.
508
509 The observed and consistent trend in the magnitude of both broad-sense and SNP-based heritability
510 estimates further emphasizes the significant contribution of additive genetic factors in the

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

511 expression of some of these traits. The heritability estimates recorded in our study also give an
512 indication of good repeatability and reproducibility of the experimental procedures. The
513 heritability estimates we found are comparable to those previously reported in other studies for
514 these traits (Oliveira et al. 2014; Oliveira et al. 2015; Njoku et al. 2015; Silva et al. 2016; Favour
515 et al. 2017; Rao et al. 2018).
516
517 Accounting for population structure and genetic relatedness in GWAS is necessary to reduce false
518 positives. PCA did not reveal the presence of substantial population stratification in our GWAS
519 panel. This is not surprising since extensive inter-crosses are routinely carried out as part of the
520 generation of new genetic variation in the IITA Cassava Breeding program. Moreover, individuals
521 from C1 and C2 largely overlapped with each other and also with the founder population (GG).
522 GWAS analysis using PCA and the kinship matrix gave very similar results to the analysis that
523 considered kinship alone in controlling for spurious associations. For this reason, as well as for
524 computational efficiency, we used the MLM model for the full analysis.
525
526 Our data replicated the previously identified associations for CMD and CGM resistance traits, dry
527 matter and carotenoid content, and also showed stronger evidence of major gene effect in this
528 larger sample. In addition to the confirmed loci, we uncovered additional genomic regions with
529 significant associations that were previously not reported. For example, two additional regions on
530 chromosome 14 were found to contribute to increased resistance to CMD virus and present useful
531 targets for further genetic analysis. Genotypes carrying favourable resistance alleles at these as
532 well as the CMD2 major effect locus on chromosome 12 exhibit high levels of resistance against
533 CMD and rarely show symptoms. However, despite the concerted global efforts to identify the
534 causal gene underlying the CMD2 locus in cassava, there has not been any breakthrough in cloning
535 and functionally identifying the actual gene. We hope that these resources will guide the cassava
536 community in narrowing down on the candidate genes to carry out the functional analysis.
537
538 Although our main interest was in traits that are under active breeding selection, we also measured
539 a number of morphological traits that we knew to be heritable. Among the morphological traits
540 presented here, to our best knowledge, only leaf shape and apical leaf pubescence were previously
541 reported (Ezenwaka et al. 2018; Zhang et al. 2018). While our study confirmed major locus for
542 apical pubescence locus on chromosome 8 we did not replicate the results of Zhang et al. (2018)
543 for leaf aspect ratio suggesting a possibility of different genetic factors.
544

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

545 Many traits of interest to crop improvement are positively or negatively correlated. Such
546 correlations can cause unfavourable changes in traits that are important but that are not under direct
547 selection. Alternatively, one trait can be used to indirectly select for another positively correlated
548 trait which is more difficult to phenotype. Genetic correlations among traits can arise due to linkage
549 disequilibrium or pleiotropy (a single gene having multiple otherwise unrelated biological effects,
550 or shared regulation of multiple genes) (Chen and Lübberstedt 2010). Correlations due to linkage
551 disequilibrium tend to be temporary and are generally considered to be less important than
552 pleiotropy. Of particular importance is the negative correlation observed between total carotenoid
553 content variation and dry matter content which confirms previous findings (Ceballos et al. 2013;
554 Njoku et al. 2015; Esuma et al. 2016; Ceballos et al. 2017; Rabbi et al. 2017; Okeke et al. 2018).
555 Both genetic linkage and pleiotropy are plausible reasons for the inverse relationship. The linkage
556 basis is supported by the co-location of major QTLs for these traits on chromosome 1 and the
557 presence of major genes in the biochemical pathways in close proximity. On the other hand,
558 genetically engineered cassava to produce and accumulate carotenoids in storage roots was found
559 in one study to have reduced dry matter content which indicates a pleiotropic effect (Beyene et al.
560 2017). Besides the region on chromosome 1, we identified additional loci on chromosomes 5, 8,
561 15, and 16 that additively explain more than 60% of the phenotypic variation for total carotenoids.
562 These and other minor effect loci could explain the lack of inverse relationship between
563 carotenoids and dry matter content in other populations (Ceballos et al. 2013). Likewise, we also
564 found several significant regions associated with dry matter content although they only explain
565 16% of the trait variation. Other notable negative but the favourable correlation in the study
566 population was between apical pubescence and CGM severity. Generally, genotypes with glabrous
567 apical leaves are more affected by the pest than pubescent ones (Raji et al. 2008; Chalwe et al.
568 2015; Ezenwaka et al. 2018). For both traits, we identified a common major locus on chromosome
569 8 that is associated with variation in the degree of pubescence as well as CGM damage severity.
570 We also found other loci that were unique to either trait suggesting additional factors may be
571 contributing to the resistance against CGM besides the presence of trichomes. Significant genetic
572 and phenotypic correlations were not observed between harvest index, dry matter content and plant
573 type, implying that they are amenable to concurrent improvement.
574

575 **Opportunities and implications**

576 The present study is based predominantly on germplasm developed by the IITA breeding program
577 and therefore represents a subset of the available diversity worldwide. While some traits such as
578 yield and yield components are universally considered, other traits, especially biotic and abiotic

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

579 stresses, are region-specific. Moreover, restricted germplasm movement due to quarantine
580 regulations makes it nearly impossible to evaluate the same population in multiple regions
581 simultaneously. Further studies using germplasm from other regions including east Africa and
582 Latin America – the centre of origin and diversity of cassava – is expected to reveal region and/or
583 population-specific large effect alleles. Such efforts are expected to enrich the catalogue of major
584 effect loci available for molecular breeding.
585
586 A major objective of this study was to provide breeders with a catalogue of major loci for marker-
587 assisted selection. Many of the previous QTL studies using bi-parental mapping populations in
588 cassava have had limited value due to low marker densities and poor genetic resolution (Ferguson
589 et al. 2012; Ceballos et al. 2015; Hershey 2017). Access to high-density genome-wide SNP
590 markers through GBS coupled with GWAS mapping approach has resulted in higher mapping
591 resolution, to within a few Kb known candidate genes for several traits in the present study. If
592 converted to allele-specific high-throughput SNP assays, SNPs tagging major loci from the present
593 study can be used to screen and identify individuals carrying favourable alleles during early stages
594 of selection. However, further validation of these loci is required to ensure they are effective across
595 environments (genotype-by-environment interaction) and populations (different genetic
596 background) before large-scale deployment. For other traits such as harvest index, the identified
597 loci only contributed to a small proportion of trait variation suggesting additional genes with small
598 effects and thus are more likely to benefit from genomic selection (GS) to estimate breeding values
599 from genome-wide marker data. GS relies on high-density marker coverage such that each QTL is
600 likely to be in linkage disequilibrium (LD) with at least one marker (Goddard and Hayes 2007). A
601 tandem approach incorporating MAS and GS is likely to increase breeding efficiency, reduce
602 breeding cycle and cost (Zhao et al. 2014). MAS can be used to screen large number of individuals
603 at seedling nursery stages to cull accessions that do not carry favourable alleles. Reduced number
604 of lines can then be genotyped at higher density for GS and allocated to field testing plots for
605 evaluation and selection for complex traits.
606

607 **Conclusions**

608 In this study, we demonstrate that the use of a diverse association mapping panel consisting of
609 landraces and improved cultivars from multiple breeding initiatives can identify SNP variants
610 associated with agronomically essential traits in cassava. The power of this study to discover
611 additional markers associated with measured traits was derived from extensive multi-locational
612 testing over four years at four IITA testing locations in Nigeria. The SNP markers we identified

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

613 provide a useful reference catalogue not only for cassava breeding programs but also studies aimed
614 at uncovering unique biological pathways necessary for advancing genetic transformation studies.
615 To realize the full potential of these marker-trait associations in population improvement, a critical
616 next step is to validate these loci in independent populations before their deployment for routine
617 use in marker-assisted selection.
618

619 **Acknowledgements**

620 This work was funded by The Bill & Melinda Gates Foundation and the Department for
621 International Development of the United Kingdom (award number OPP1048542) through the
622 “Next Generation Cassava Breeding Project”; and the CGIAR-Research Program on Roots, Tubers
623 and Bananas. Thanks to the technical staff at IITA Cassava Breeding Program for implementing
624 the field experiments.
625

626 **Authors contributions**

627 IYR, CE, JLJ, and PK conceived and designed the study; IYR, SIK, GB, AA, and MY performed
628 analyses and wrote the manuscript; CE, EL, EP, MW, JLJ, and PK edited the manuscript; CA, KO,
629 RU, ASI, and PP Implemented field trials, generated and curated data; and PK Provided overall
630 coordination and leadership
631

632 **Competing interests**

633 The authors declare no competing interests.
634

635 **Data availability**

636 The data that support the findings of this study are openly available in CassavaBase at
637 ftp://ftp.cassavabase.org/manuscripts/PlantMolBiol_Rabbi_et_al_2020/
638
639

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

640 **References**

641 Abass AB, Towo E, Mukuka I, Okechukwu R, Ranaivoson R, Tarawali G, Kanju E (2014)
642 Growing cassava: a training manual from production to postharvest

643 Akano A, Dixon A, Mba C, Barrera E, Fregene M (2002) Genetic mapping of a dominant gene
644 conferring resistance to cassava mosaic disease. Theor Appl Genet 105:521–525 .
645 https://doi.org/10.1007/s00122-002-0891-7

646 Alabi OJ, Mulenga RM, Legg JP (2015) Cassava mosaic: Virus diseases of tropical and
647 subtropical crops. Virus Dis Trop Subtrop Crops 42–55

648 Allan AC, Hellens RP, Laing WA (2008) MYB transcription factors that colour our fruit. Trends
649 Plant Sci 13:99–102

650 Atser G, Dixon A, Ekeleme F, Chikoye D, Dashiell KE, Ayankanmi T, Hauser S, Agada M,
651 Okwusi M, Sokoya G (2017) The ABC of weed management in cassava production in
652 Nigeria: a training manual

653 Bates D, Maechler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using
654 lme4. J Stat Softw 67:1–48 . https://doi.org/10.18637/jss.v067.i01

655 Batra S, Carlson J, Hayes R, Shu S, Schmutz J, Rokhsar D (2014) Phytozome Comparative Plant
656 Genomics Portal. 1–2

657 Bechoff A, Tomlins K, Westby A, Fliedel G, Hershey C, Dufour D, Becerra Lopez-lavalle LA
658 (2018) Cassava traits and end-user preference: Relating traits to consumer liking, sensory
659 perception, and genetics. Crit Rev Food Sci Nutr

660 Beyene G, Chauhan RD, Ilyas M, Wagaba H, Fauquet CM, Miano D, Alicai T, Taylor NJ (2017)
661 A Virus-Derived Stacked RNAi Construct Confers Robust Resistance to Cassava Brown
662 Streak Disease. Front Plant Sci 7: . https://doi.org/10.3389/fpls.2016.02052

663 Browning BL (2016) Beagle 4.1

664 Burns A, Gleadow R, Cliff J, Zacarias A, Cavagnaro T (2010) Cassava: The Drought, War and
665 Famine Crop in a Changing World. Sustainability 2:3572–3607 .
666 https://doi.org/10.3390/su2113572

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

667 Byrne DH, Guerrero JM, Belloti AC, Gracen VE (1982) Behavior and Development of
668 Mononychellus tanajoa (Acari: Tetranychidae) on Resistant and Susceptible Cultivars of
669 Cassava1. J Econ Entomol 75:924–927 . https://doi.org/10.1093/jee/75.5.924

670 CABI (2019) Invasive Species Compendium. https://www.cabi.org/isc. Accessed 11 Dec 2019

671 Ceballos H, Davrieux F, Talsma EF, Belalcazar J, Chavarriaga P, Andersson MS (2017)
672 Carotenoids in Cassava Roots. In: Carotenoids. InTech

673 Ceballos H, Kawuki RS, Gracen VE, Yencho GC, Hershey CH (2015) Conventional breeding,
674 marker-assisted selection, genomic selection and inbreeding in clonally propagated crops:
675 a case study for cassava. Theor Appl Genet. https://doi.org/10.1007/s00122-015-2555-4

676 Ceballos H, Kulakow P, Hershey C (2012) Cassava Breeding: Current Status, Bottlenecks and
677 the Potential of Biotechnology Tools. Trop Plant Biol 5:73–87 .
678 https://doi.org/10.1007/s12042-012-9094-9

679 Ceballos H, Morante N, Sánchez T, Ortiz D, Aragón I, Chávez AL, Pizarro M, Calle F, Dufour
680 D (2013) Rapid cycling recurrent selection for increased carotenoids content in cassava
681 roots. Crop Sci 53: . https://doi.org/10.2135/cropsci2013.02.0123

682 Chalwe A, Melis R, Shanahan P, Chiona M (2015) Inheritance of resistance to cassava green
683 mite and other useful agronomic traits in cassava grown in Zambia. Euphytica 205:103–
684 119 . https://doi.org/10.1007/s10681-015-1404-5

685 Chen H, Wang C, Conomos MP, Stilp AM, Li Z, Sofer T, Szpiro AA, Chen W, Brehm JM,
686 Celedón JC, Redline S, Papanicolaou GJ, Thornton TA, Laurie CC, Rice K, Lin X (2016)
687 Control for Population Structure and Relatedness for Binary Traits in Genetic
688 Association Studies via Logistic Mixed Models. Am J Hum Genet 98:653–666 .
689 https://doi.org/10.1016/j.ajhg.2016.02.012

690 Chen Y, Lübberstedt T (2010) Molecular basis of trait correlations. Trends Plant Sci 15:454–461

691 Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA minipreparation: Version II. Plant Mol
692 Biol Report 1:19–21 . https://doi.org/10.1007/BF02712670

693 Dixon AGO, Ssemakula G (2008) Prospects for cassava breeding in Sub-Saharan Africa in the
694 next decade. J Food Agric Environ 6:256–262

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

695 Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A
696 robust, simple genotyping-by-sequencing (GBS) approach for high diversity species.
697 PLoS ONE 6:1–10 . https://doi.org/10.1371/journal.pone.0019379

698 Esuma W, Herselman L, Labuschagne MT, Ramu P, Lu F, Baguma Y, Buckler ES, Kawuki RS
699 (2016) Genome-wide association mapping of provitamin A carotenoid content in cassava.
700 Euphytica 212:97–110 . https://doi.org/10.1007/s10681-016-1772-5

701 Ezenwaka L, Del Carpio Dunia P, Jannink J-L, Rabbi I, Danquah E, Asante I, Danquah A, Blay
702 E, Egesi C (2018) Genome-wide association study of resistance to cassava green mite
703 pest and related traits in cassava. Crop Sci 58:1907–1918

704 Favour E, Emeka N, Chiedozie E, Bunmi O, Emmanuel O (2017) Genetic variability, heritability
705 and variance components of some yield and yield related traits in second backcross
706 population (BC2) of cassava. Afr J Plant Sci 11:185–189 .
707 https://doi.org/10.5897/AJPS2015.1324

708 Ferguson M, Rabbi I, Kim DJ, Gedil M, Lopez-Lavalle LAB, Okogbenin E (2012) Molecular
709 Markers and Their Application to Cassava Breeding: Past, Present and Future. Trop Plant
710 Biol 5:95–109 . https://doi.org/10.1007/s12042-011-9087-0

711 Fukuda WMG, Guevara CL, Kawuki R, Ferguson ME (2010) Selected morphological and
712 agronomic descriptors for the characterization of cassava. Int Inst Trop Agric 19–19

713 Furumizu C, Alvarez JP, Sakakibara K, Bowman JL (2015) Antagonistic Roles for KNOX1 and
714 KNOX2 Genes in Patterning the Land Plant Body Plan Following an Ancient Gene
715 Duplication. PLOS Genet 11:e1004980 . https://doi.org/10.1371/journal.pgen.1004980

716 García-Ruiz A, Cole JB, VanRaden PM, Wiggans GR, Ruiz-López FJ, Van Tassell CP (2016)
717 Changes in genetic selection differentials and generation intervals in US Holstein dairy
718 cattle as a result of genomic selection. Proc Natl Acad Sci 113:E3995 .
719 https://doi.org/10.1073/pnas.1519061113

720 Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-
721 GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS ONE 9:e90346–
722 e90346 . https://doi.org/10.1371/journal.pone.0090346

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

723 Goddard ME, Hayes BJ (2007) Genomic selection. J Anim Breed Genet Z Tierzuchtung
724 Zuchtungsbiologie 124:323–330 . https://doi.org/10.1111/j.1439-0388.2007.00702.x

725 Gonçalves L, Rodrigues R, Diz M, Robaina R, Júnior A, Carvalho A, Gomes V (2013)
726 Peroxidase is involved in Pepper yellow mosaic virus resistance in Capsicum baccatum
727 var. pendulum. Genet Mol Res 12:1411–1420

728 Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten
729 U, Putnam N (2011) Phytozome: a comparative platform for green plant genomics.
730 Nucleic Acids Res 40:D1178–D1186

731 Gould K, Davies KM, Winefield C (2008) Anthocyanins: biosynthesis, functions, and
732 applications. Springer Science & Business Media

733 Hamblin MT, Rabbi IY (2014) The Effects of Restriction-Enzyme Choice on Properties of
734 Genotyping-by-Sequencing Libraries: A Study in Cassava (). Crop Sci 0:0–0 .
735 https://doi.org/10.2135/cropsci2014.02.0160

736 Haouazine-Takvorian N, Tymowska-Lalanne Z, Takvorian A, Tregear J, Lejeune B, Lecharny
737 A, Kreis M (1997) Characterization of two members of the Arabidopsis thaliana gene
738 family, Atβfruct3 and Atβfruct4, coding for vacuolar invertases. Gene 197:239–251

739 Hershey CH (2017) Marker-assisted selection in cassava breeding Ismail Y. Rabbi, International
740 Institute of Tropical Agriculture (IITA), Nigeria. In: Achieving sustainable cultivation of
741 cassava Volume 2. Burleigh Dodds Science Publishing, pp 121–127

742 Himi E, Noda K (2005) Red grain colour gene (R) of wheat is a Myb-type transcription factor.
743 Euphytica 143:239–242

744 Ikeogu UN, Akdemir D, Wolfe MD, Okeke UG, Amaefula C, Jannink J-L, Egesi CN (2019)
745 Genetic Correlation, Genome-Wide Association and Genomic Prediction of Portable
746 NIRS Predicted Carotenoids in Cassava Roots. Front Plant Sci 10:1570–1570

747 Jarvis A, Ramirez-Villegas J, Herrera Campo BV, Navarro-Racines C (2012) Is Cassava the
748 Answer to African Climate Change Adaptation? Trop Plant Biol 5:9–29 .
749 https://doi.org/10.1007/s12042-012-9096-7

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

750 Kawano K (2003) Thirty years of cassava breeding for productivity—biological and social
751 factors for success. Crop Sci 43:1325–1335

752 Kawano K, Daza P, Amaya A, Rios M, Goncalves WMF (1978) Evaluation of cassava
753 germplasm for productivity. Crop Sci 18:377–380

754 Kawano K, Maria W, Fuluda G, Cenpuldee U, Maria W, Fuknda G, Cenpnkdee U, Fnknda G,
755 Cenpukdee U (1987) Genetic and Environmental Effects on Dry Matter Content of
756 Cassava Root. Crop Sci 27:69–74 .
757 https://doi.org/10.2135/cropsci1987.0011183X002700010018x

758 Kawano K, Narintaraporn K, Narintaraporn P, Sarakarn S, Limsila A, Limsila J, Suparhan D,
759 Sarawat V, Watananonta W (1998) Yield improvement in a multistage breeding program
760 for cassava. Crop Sci 38:325–332 .
761 https://doi.org/10.2135/cropsci1998.0011183X003800020007x

762 Kayondo SI, Pino Del Carpio D, Lozano R, Ozimati A, Wolfe M, Baguma Y, Gracen V, Offei S,
763 Ferguson M, Kawuki R, Jannink J-L (2018) Genome-wide association mapping and
764 genomic prediction for CBSD resistance in Manihot esculenta. Sci Rep 8:1549–1549 .
765 https://doi.org/10.1038/s41598-018-19696-1

766 Kobayashi S, Ishimaru M, Hiraoka K, Honda C (2002) Myb-related genes of the Kyoho grape
767 (Vitis labruscana) regulate anthocyanin biosynthesis. Planta 215:924–933

768 Legg J, Owor B, Sseruwagi P, Ndunguru J (2006) Cassava mosaic virus disease in East and
769 Central Africa: epidemiology and management of a regional pandemic. Adv Virus Res
770 67:355–418

771 LiLin-Yin (2019) CMplot: Circle Manhattan Plot

772 Listgarten J, Lippert C, Kadie CM, Davidson RI, Eskin E, Heckerman D (2012) Improved linear
773 mixed models for genome-wide association studies. Nat Methods 9:525

774 Mitchell C, Brennan RM, Graham J, Karley AJ (2016) Plant Defense against Herbivorous Pests:
775 Exploiting Resistance and Tolerance Traits for Sustainable Crop Protection. Front Plant
776 Sci 7:1132 . https://doi.org/10.3389/fpls.2016.01132

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

777 Nägele T, Henkel S, Hörmiller I, Sauter T, Sawodny O, Ederer M, Heyer AG (2010)
778 Mathematical modeling of the central carbohydrate metabolism in Arabidopsis reveals a
779 substantial regulatory influence of vacuolar invertase on whole plant carbon metabolism.
780 Plant Physiol 153:260–272 . https://doi.org/10.1104/pp.110.154443

781 Nesi N, Jond C, Debeaujon I, Caboche M, Lepiniec L (2001) The Arabidopsis TT2 gene encodes
782 an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin
783 accumulation in developing seed. Plant Cell 13:2099–2114

784 Njoku DN, Gracen VE, Offei SK, Asante IK, Egesi CN, Kulakow P, Ceballos H (2015) Parent-
785 offspring regression analysis for total carotenoids and some agronomic traits in cassava.
786 Euphytica 206:657–666 . https://doi.org/10.1007/s10681-015-1482-4

787 Okechukwu R, Dixon AG (2008) Genetic gains from 30 years of cassava breeding in Nigeria for
788 storage root yield and disease resistance in elite cassava genotypes. J Crop Improv
789 22:181–208

790 Okeke UG, Akdemir D, Rabbi I, Kulakow P, Jannink JL (2018) Regional heritability mapping
791 provides insights into dry matter content in african white and yellow cassava populations.
792 Plant Genome 11: . https://doi.org/10.3835/plantgenome2017.06.0050

793 Oliveira EJ de, Aidar S de T, Morgante CV, Chaves AR de M, Cruz JL, Coelho Filho MA
794 (2015) Genetic parameters for drought-tolerance in cassava. Pesqui Agropecuária Bras
795 50:233–241

796 Oliveira EJ, Santana FA, Oliveira LA, Santos VS (2014) Genetic parameters and prediction of
797 genotypic values for root quality traits in cassava using REML/BLUP. Genet Mol Res
798 13:6683–6700 . https://doi.org/10.4238/2014.August.28.13

799 Oshima Y, Shikata M, Koyama T, Ohtsubo N, Mitsuda N, Ohme-Takagi M (2013) MIXTA-like
800 transcription factors and WAX INDUCER1/SHINE1 coordinately regulate cuticle
801 development in Arabidopsis and Torenia fournieri. Plant Cell 25:1609–1624

802 Owor B, Legg J, Okao-Okuja G, Obonyo R, Ogenga-Latigo M (2004) The effect of cassava
803 mosaic geminiviruses on symptom severity, growth and root yield of a cassava mosaic
804 virus disease-susceptible cultivar in Uganda. Ann Appl Biol 145:331–337

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

805 Palta JP (1990) Leaf chlorophyll content. Remote Sens Rev 5:207–213 .
806 https://doi.org/10.1080/02757259009532129

807 Parmar A, Sturm B, Hensel O (2017) Crops that feed the world: Production and improvement of
808 cassava for food, feed, and industrial uses. Food Secur 9:907–927

809 Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, Rodriguez F,
810 Fauquet C, Tohme J, Harkins T, Rokhsar DS, Rounsley S (2012) The Cassava Genome:
811 Current Progress, Future Directions. Trop Plant Biol 5:88–94 .
812 https://doi.org/10.1007/s12042-011-9088-z

813 Prudencio YC, Al-Hassan R (1994) The food security stabilization roles of cassava in Africa.
814 Food Policy 19:57–64 . https://doi.org/10.1016/0306-9192(94)90008-6

815 Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR,
816 Willer CJ (2010) LocusZoom: regional visualization of genome-wide association scan
817 results. Bioinformatics 26:2336–2337

818 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de
819 Bakker PIW, Daly MJ, Sham PC (2007) PLINK: A Tool Set for Whole-Genome
820 Association and Population-Based Linkage Analyses. Am J Hum Genet 81:559–575 .
821 https://doi.org/10.1086/519795

822 R Development Core Team (2016) R: A language and environment for statistical computing. R
823 Foundation for Statistical Computing, R Foundation for Statistical Computing, Vienna,
824 Austria

825 Rabbi IY, Hamblin MT, Kumar PL, Gedil MA, Ikpan AS, Jannink JL, Kulakow PA (2014)
826 High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using
827 genotyping-by-sequencing and its implications for breeding. Virus Res 186:87–96 .
828 https://doi.org/10.1016/j.virusres.2013.12.028

829 Rabbi IY, Udoh LI, Wolfe M, Parkes EY, Gedil MA, Dixon A, Ramu P, Jannink J-L, Kulakow P
830 (2017) Genome-Wide Association Mapping of Correlated Traits in Cassava: Dry Matter
831 and Total Carotenoid Content. Plant Genome 0:0–0 .
832 https://doi.org/10.3835/plantgenome2016.09.0094

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

833 Raji A, Ladeinde O, Dixon A (2008) Screening landraces for additional sources of field
834 resistance to cassava mosaic disease and green mite for integration into the cassava
835 improvement program. J Integr Plant Biol 50:311–318

836 Ramu P, Esuma W, Kawuki R, Rabbi IY, Egesi C, Bredeson JV, Bart RS, Verma J, Buckler ES,
837 Lu F (2017) Cassava haplotype map highlights fixation of deleterious mutations during
838 clonal propagation. Nat Genet 49:959–963 . https://doi.org/10.1038/ng.3845

839 Rao BB, Swami D, Ashok P, Babu BK, Ramajayam D, Sasikala K (2018) Estimation of Genetic
840 Variability and Heritability for Yield and Its Related Components in Cassava (Manihot
841 esculenta Crantz) Genotypes. Int J Curr Microbiol App Sci 7:287–297

842 Rentería ME, Cortes A, Medland SE (2013) Using PLINK for genome-wide association studies
843 (GWAS) and data analysis. In: Gondro C, van der Werf J, Hayes B (eds) Methods in
844 Molecular Biology. Humana Press, Totowa, pp 193–213

845 Saltzman A, Birol E, Bouis HE, Boy E, De Moura FF, Islam Y, Pfeiffer WH (2013)
846 Biofortification: Progress toward a more nourishing future. Glob Food Secur 2:9–17 .
847 https://doi.org/10.1016/j.gfs.2012.12.003

848 Sánchez T, Ceballos H, Dufour D, Ortiz D, Morante N, Calle F, Zum Felde T, Dominguez M,
849 Davrieux F (2014) Prediction of carotenoids, cyanide and dry matter contents in fresh
850 cassava root using NIRS and Hunter color techniques. Food Chem 151:444–451

851 Shukla P (1976) Preliminary report on the green mite (Mononychellus tanajoa, Bonder)
852 resistance in Tanzanian local cassava varieties. East Afr Agric For J 42:55–59

853 Silva R de S, Moura EF, Neto JT de F, Sampaio JE (2016) Genetic parameters and agronomic
854 evaluation of cassava genotypes. Pesqui Agropecu Bras 51:834–841 .
855 https://doi.org/10.1590/S0100-204X2016000700006

856 Sinclair TR (1998) Historical changes in harvest index and crop nitrogen accumulation. Crop Sci
857 38:638–643

858 Speed D, Hemani G, Johnson MR, Balding DJ (2012) Improved heritability estimation from
859 genome-wide SNPs. Am J Hum Genet 91:1011–1021

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

860 Takenaka A (1994) Effects of leaf blade narrowness and petiole length on the light capture
861 efficiency of a shoot. Ecol Res 9:109–114

862 Thottappilly G, Thresh JM, Calvert LA, Winter S (2003) Cassava. In: Loebenstein G,
863 Thottappilly G (eds) Virus and Virus-like Diseases of Major Crops in Developing
864 Countries. Springer Netherlands, Dordrecht, pp 107–165

865 Varshney RK, Terauchi R, McCouch SR (2014) Harvesting the Promising Fruits of Genomics:
866 Applying Genome Sequencing Technologies to Crop Breeding. PLOS Biol 12:e1001883
867 . https://doi.org/10.1371/journal.pbio.1001883

868 Wei T, Simko V (2017) R package "corrplot": Visualization of a Correlation Matrix (Version
869 0.84). URL Httpsgithub Comtaiyuncorrplot

870 Welsch R, Arango J, Bär C, Salazar B, Al-Babili S, Beltrán J, Chavarriaga P, Ceballos H, Tohme
871 J, Beyer P (2010) Provitamin A Accumulation in Cassava (Manihot
872 esculenta) Roots Driven by a Single Nucleotide Polymorphism in a Phytoene
873 Synthase Gene. Plant Cell 22:3348 . https://doi.org/10.1105/tpc.110.077560

874 Wolfe MD, Rabbi IY, Egesi C, Hamblin M, Kawuki R, Kulakow P, Lozano R, Carpio DPD,
875 Ramu P, Jannink J-L (2016) Genome-Wide Association and Prediction Reveals Genetic
876 Architecture of Cassava Mosaic Disease Resistance and Prospects for Rapid Genetic
877 Improvement. Plant Genome 9:0–0 . https://doi.org/10.3835/plantgenome2015.11.0118

878 Wu G, Shortt BJ, Lawrence EB, Leon J, Fitzsimmons KC, Levine EB, Raskin I, Shah DM
879 (1997) Activation of host defense mechanisms by elevated production of H2O2 in
880 transgenic plants. Plant Physiol 115:427–435

881 Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: A tool for genome-wide complex
882 trait analysis. Am J Hum Genet 88:76–82 . https://doi.org/10.1016/j.ajhg.2010.11.011

883 Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL (2014) Advantages and pitfalls in the
884 application of mixed-model association methods. Nat Genet 46:100–106 .
885 https://doi.org/10.1038/ng.2876

886 Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM (2017) Concepts, estimation and
887 interpretation of SNP-based heritability. Nat Genet 49:1304

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

888 Ye X, Pan S, Kuc J (1990) Activity, isozyme pattern, and cellular localization of peroxidase as
889 related to systemic resistance of tobacco to blue mold(Peronospora tabacina) and to
890 tobacco mosaic virus. Phytopathology 80:1295–1299

891 Zhang S, Chen X, Lu C, Ye J, Zou M, Lu K, Feng S, Pei J, Liu C, Zhou X (2018) Genome-wide
892 association studies of 11 agronomic traits in cassava (Manihot esculenta Crantz). Front
893 Plant Sci 9:503–503

894 Zhao Y, Mette MF, Gowda M, Longin CFH, Reif JC (2014) Bridging the gap between marker-
895 assisted and genomic selection of heading time and plant height in hybrid wheat.
896 Heredity 112:638–645 . https://doi.org/10.1038/hdy.2014.1

897

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

898 **Tables**

899 **Table 1** Name, brief description, and classification of the 14 traits assessed in the present study

900 **Table 2** Broad-sense and SNP heritability estimates, variance components and coefficients of
901 variation of 14 traits in cassava GWAS panel

902 **Table 3** Summary of most significant SNP markers at each major trait linked locus for the 14
903 studied traits

904

905

906 **Figure captions**

907 **Figure 1** Heatmap of pairwise trait correlations using BLUPS from the 14 traits

908 **Figure 2** Overview of SNP genotyping data. (a) The density of SNPs on the 18 chromosomes of
909 the cassava association mapping panel within 100 Kb window. (b) Histogram of minor allele
910 frequency distribution. (c) Genome-wide Linkage disequilibrium (LD) decay for the cassava
911 accessions in the panel showing the squared correlations (r^2) between markers by marker physical
912 distance (kb). The blue smoothening curve (LOESS) and the average LD were fitted to the LD
913 decay.

914 **Figure 3** An assessment of population structure based on Principal Components Analysis of
915 101,521 SNP marker data (MAF>1%) for 5,130 individual cassava clones. (a) Scatter plot of the
916 first two PCs; (b) the proportion of genetic variation explained by the first ten 10 PCs; and (c)
917 Heatmap showing pairwise Kinship matrix

918 **Figure 4** Manhattan plots for GWAS for 14 traits of 5,130 cassava accessions using MLM analysis
919 approach. A total of 101,521 SNP markers were used for the GWAS analyses with the red
920 horizontal line representing Bonferroni adjusted genome-wide significance threshold
921 (α=0.05/101521=4.93 x 10^{-7}). The QQ-plots inset - right with observed p-values on the y-axis and
922 expected p-values on the x-axis

923 **Figure 5** Multiple Marker trait regression barplot across traits and the proportion of phenotypic
924 variance explained. Numbers above bar plot denote the number of loci in the regression model

925

926

927

928 **Supplementary Tables and Figures**

929 **Supplementary Table S1** Clone names, GBS ID, pedigree, clonal generation and BLUPS for the
930 14 traits.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

931 **Supplementary Table S2** The pairwise estimates of genetic (lower diagonal) and phenotypic
932 correlations (upper diagonal) using GBLUPS and BLUPS respectively among the 14 traits.
933 **Supplementary Table S3** Summary information of the potential candidate genes identified in the
934 vicinity of the GWAS hits for analyzed traits.
935
936 **Supplementary Figure S1** Variation and trends in phenotypic data for 14 morphological,
937 agronomic, quality and defensive traits in a diverse cassava association mapping panel evaluated
938 between 2013 to 2016 across four locations. Histograms of 14 traits were based on de-regressed
939 BLUPs distributions measured on 5,130 cassava accessions.
940
941 **Supplementary Figure S2** Manhattan plots for GWAS for 14 traits of 5,130 cassava accessions
942 using the leave-one-chromosome-out "MLMe" analysis approach. A total of 101,521 SNP markers
943 were used for that GWAS analyses with the red horizontal line representing genome-wide
944 significance threshold (5%). The QQ-plots inset - right with observed p-values on the y-axis and
945 expected p-values on the x-axis.
946
947 **Supplementary Figure S3** Allelic substitution effects in the most significant SNP at each locus
948 identified for each of the 14 studied traits. Trait BLUPs distribution on y-axis and SNP genotype status of
949 the marker on x-axis. SNPs were converted to dosage format, where 0, 1, and 2 indicates the copies of the
950 minor alleles. The first allele in the suffix of a SNP name denotes the allele being counted in the dosage
951 coding. For example, dosage score of 2 in SNP S12_7926132_G(T) means homozygous for "G", a score
952 of 1 means heterozygote "GT", and a score of 0 means homozygote alternate allele "T". (A) CMD severity;
953 (B) CGM severity; (C) Apical leaf pubescence; (D) Leaf shape(E); Apical leaf colour; (F) Mature leaf
954 greenness; (G) Petiole colour; (H) Harvest index; (I) Plant type; (J) Outer stem colour; (K) Total carotenoids
955 content; (L) Dry matter content; (M) Periderm colour; (N) Root cortex colour.
956
957
958 **Supplementary Figure S4** Regional Manhattan plots for each locus-trait combination. Plots
959 include candidate genes within 100 Kb of the top SNP marker (50 kb upstream, 50 kb downstream)
960 with some adjustments based on the extent of local linkage disequilibrium with the candidate
961 marker. SNPs are colour coded based on linkage disequilibrium with the top marker. (A) CMD
962 severity; (B) CGM severity; (C) Apical leaf pubescence; (D) Leaf shape; (E) Apical leaf colour; (F) Mature
963 leaf greenness; (G) Petiole colour; (H) Harvest index; (I) Plant type; (J) Outer stem colour; (K) Total
964 carotenoids content; (L) Dry matter content; (M) Periderm colour; (N) Root cortex colour.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

965 **Table 1** Name, brief description, and classification of the 14 traits assessed in the present study

Name of Trait	**Description**	**Class**
Cassava mosaic disease (CMD) severity	The visual rating of symptoms caused by cassava mosaic virus	Biotic stress
Cassava green mite (CGM) severity	The visual rating of damage caused by cassava green mite	Biotic stress
Apical pubescence	Pubescence of young apical leaves	Morphological
Leaf shape	The shape of central leaf taken from a mid-height position.	Morphological
Apical leaf colour	Colour of unexpanded apical leaves	Morphological
Mature leaf greenness	Colour of first fully expanded leaf, an indicator of leaf chlorophyll content	Morphological
Petiole colour	Pigmentation of leaf petioles	Morphological
Harvest index	Harvest index	Agronomic traits
Plant type	Plant architecture on a 1-5 scale	Morphological
Outer stem colour	Stem colour nine months after planting	Morphological
Total carotenoid (colour chart)	Level of yellowness in cassava storage root pulp (parenchyma) due to variation in carotenoid content	Quality traits
Dry matter content	Percentage of dry matter content of storage roots	Quality traits
Storage root periderm colour	Colour of the outer surface of storage root periderm (outer skin)	Morphological
Storage root cortex colour	Colour of the outer surface of storage root cortex (inner skin)	Morphological

966
967

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

968 **Table 2** Broad-sense and SNP heritability estimates, variance components and coefficients of
969 variation of 14 traits in cassava GWAS panel

Trait	SNP-h_2	H_2	σ_g	$\sigma_{g\times e}$	σ_e	CV
CMD severity	0.434	0.776	0.783	0.086	0.140	63
CGM severity	0.165	0.149	0.074	0.177	0.244	19
Apical pubescence	0.502	0.531	0.175	0.083	0.071	113
Leaf shape	0.499	0.510	0.679	0.475	0.178	22
Apical leaf colour	0.496	0.567	1.610	0.629	0.601	23
Mature leaf greenness	0.568	0.531	0.427	0.186	0.191	18
Petiole colour	0.716	0.754	3.322	0.602	0.485	33
Harvest index	0.308	0.538	0.010	0.002	0.007	28
Plant type	0.384	0.369	0.376	0.180	0.465	38
Outer stem colour	0.516	0.388	0.907	0.191	1.241	22
Total carotenoids content (colour chart)	0.675	0.726	0.401	0.066	0.085	49
Dry matter content	0.565	0.521	14.776	3.385	10.184	15
Root periderm colour	0.548	0.610	0.190	0.035	0.086	19
Root cortex colour	0.518	0.415	0.070	0.036	0.062	30

970 Where: H_2 is the Broad-sense heritability, σ_g is the clonal genotypic variance, $\sigma_{g\times e}$ is the variance due to genotype by
971 environment (G×E), and σ_e being the residual variance.

972

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

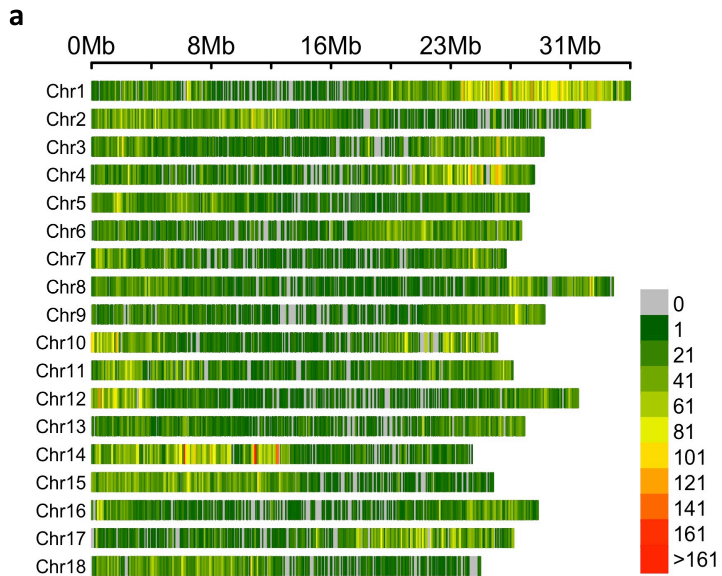
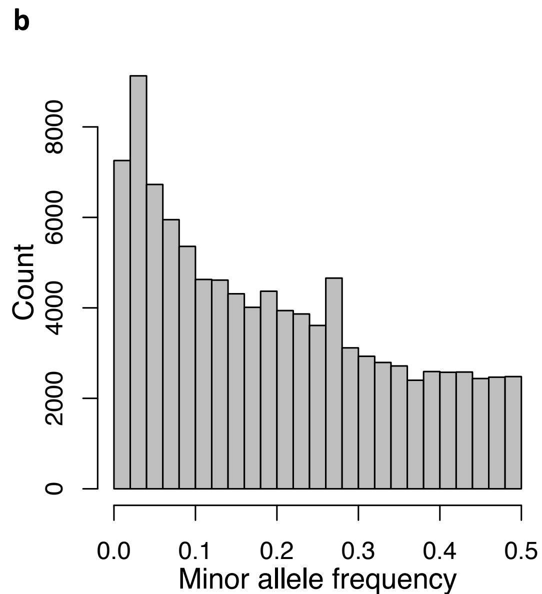
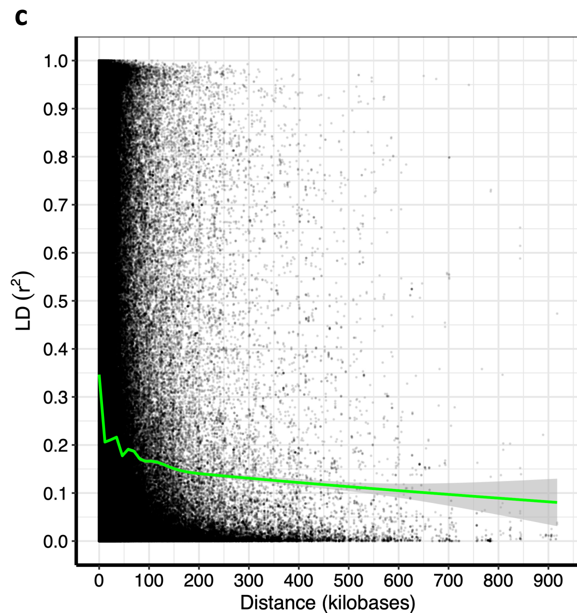
Table 3 Summary of most significant SNP markers at each major trait linked locus for the 14 traits

Trait	SNP	Minor Allele	Major Allele	MAF	beta	se	P-value	beta*	se*	P-value*
CMD severity	S12_7926132	G	T	0.44	0.82	0.04	**1.7×10^{-112}**	0.89	0.02	**P ≈ 0.0**
CMD severity	S14_4626854	A	G	0.40	-0.23	0.03	**1.7×10^{-14}**	-0.23	0.03	**1.0×10^{-14}**
CMD severity	S14_9004550	T	C	0.17	0.28	0.03	**4.2×10^{-17}**	0.27	0.03	**1.7×10^{-16}**
CGM severity	S1_28672656	A	T	0.28	0.09	0.04	1.4×10^{-2}	0.10	0.02	**1.1×10^{-10}**
CGM severity	S8_6409580	C	G	0.41	-0.18	0.02	**3.6×10^{-25}**	-0.18	0.01	**4.7×10^{-37}**
CGM severity	S12_5524524	C	T	0.38	-0.07	0.02	6.3×10^{-5}	-0.07	0.01	**1.3×10^{-7}**
CGM severity	S17_23749968	A	G	0.31	0.09	0.02	1.5×10^{-5}	0.08	0.01	**1.1×10^{-7}**
Apical pubescence	S8_6409580	C	G	0.41	0.41	0.02	**1.1×10^{-154}**	0.42	0.01	**4.3×10^{-293}**
Apical pubescence	S9_1588034	A	G	0.47	0.08	0.02	**4.2×10^{-7}**	0.08	0.01	**2.0×10^{-14}**
Apical pubescence	S11_5727254	A	G	0.09	-0.19	0.03	**1.9×10^{-12}**	-0.19	0.03	**1.8×10^{-12}**
Apical pubescence	S12_5524524	C	T	0.38	0.07	0.02	2.3×10^{-4}	0.09	0.01	**6.8×10^{-18}**
Apical pubescence	S16_1501762	G	A	0.30	-0.06	0.02	2.8×10^{-3}	-0.06	0.01	**8.1×10^{-8}**
Leaf shape	S15_10273255	A	G	0.03	2.70	0.10	**3.7×10^{-174}**	2.90	0.08	**P ≈ 0.0**
Leaf shape	S15_20573383	G	C	0.02	0.73	0.11	**1.8×10^{-11}**	1.47	0.10	**1.6×10^{-52}**
Apical leaf colour	S2_6086714	A	T	0.43	-1.22	0.06	**6.1×10^{-83}**	-1.26	0.04	**5.7×10^{-267}**
Apical leaf colour	S3_4745233	G	A	0.23	-0.35	0.06	**4.4×10^{-9}**	-0.34	0.04	**1.7×10^{-18}**
Apical leaf colour	S8_6061421	G	C	0.41	-0.32	0.05	**1.9×10^{-11}**	-0.27	0.03	**2.4×10^{-16}**
Mature leaf greenness	S1_23452638	T	A	0.16	-1.11	0.04	**2.7×10^{-167}**	-1.24	0.03	**P ≈ 0.0**
Petiole colour	S1_23452638	T	A	0.16	2.73	0.10	**9.8×10^{-180}**	2.73	0.10	**9.8×10^{-180}**
Harvest index	S2_2809137	G	T	0.09	-0.04	0.01	**3.0×10^{-8}**	-0.04	0.01	**5.6×10^{-15}**
Harvest index	S12_6055806	A	G	0.24	-0.03	0.00	**5.4×10^{-24}**	-0.02	0.00	**2.0×10^{-17}**
Plant type	S1_3192405	T	C	0.28	0.20	0.05	1.7×10^{-5}	0.22	0.04	**3.8×10^{-8}**
Plant type	S1_25303195	C	A	0.48	0.23	0.05	3.4×10^{-5}	0.27	0.04	**5.3×10^{-14}**
Outer stem colour	S2_13928566	C	G	0.39	-0.44	0.10	**3.6×10^{-14}**	-0.45	0.06	**6.5×10^{-15}**
Outer stem colour	S8_13604799	G	A	0.34	-1.06	0.06	**8.3×10^{-69}**	-1.06	0.05	**8.7×10^{-89}**
Outer stem colour	S8_22630799	G	A	0.25	-0.58	0.07	3.3×10^{-17}	-0.76	0.06	**1.3×10^{-37}**
Total carotenoids (color chart)	S1_24159583	T	C	0.30	0.27	0.02	**5.3×10^{-57}**	0.37	0.01	**1.3×10^{-219}**
Total carotenoids (color chart)	S1_24636113	G	A	0.23	0.49	0.03	**1.8×10^{-78}**	0.57	0.02	**1.3×10^{-270}**
Total carotenoids (color chart)	S1_30543962	G	A	0.10	0.18	0.03	**6.3×10^{-8}**	0.40	0.02	**2.4×10^{-76}**
Total carotenoids (color chart)	S5_3387558	T	C	0.10	0.21	0.02	**5.5×10^{-17}**	0.20	0.02	**2.0×10^{-16}**
Total carotenoids (color chart)	S8_4319215	A	C	0.01	0.22	0.05	5.2×10^{-5}	0.23	0.04	**3.4×10^{-8}**
Total carotenoids (color chart)	S8_25598183	T	G	0.03	0.18	0.04	6.3×10^{-6}	0.18	0.03	**1.1×10^{-8}**
Total carotenoids (color chart)	S15_7659426	G	T	0.26	-0.07	0.02	2.0×10^{-3}	-0.06	0.01	**3.2×10^{-7}**
Total carotenoids (color chart)	S16_484011	G	T	0.35	0.09	0.03	7.6×10^{-4}	0.05	0.01	**8.9×10^{-8}**
Dry matter content	S1_24636113	G	A	0.23	-1.32	0.24	**5.0×10^{-8}**	-1.68	0.14	**1.7×10^{-33}**
Dry matter content	S6_20589894	G	A	0.48	0.85	0.17	**3.4×10^{-7}**	0.78	0.09	**1.7×10^{-16}**
Dry matter content	S12_5524524	C	T	0.37	0.68	0.17	9.8×10^{-5}	0.69	0.10	**8.0×10^{-12}**
Dry matter content	S15_1012346	C	T	0.44	-0.95	0.20	2.2×10^{-6}	-0.84	0.10	**4.0×10^{-17}**
Dry matter content	S16_25663808	T	C	0.35	-0.48	0.18	6.8×10^{-3}	-0.69	0.10	**4.2×10^{-12}**
Periderm colour	S3_4545411	G	C	0.43	-0.38	0.02	**1.4×10^{-123}**	-0.42	0.01	**7.3×10^{-288}**
Cortex colour	S1_3047840	T	G	0.01	1.08	0.05	**1.6×10^{-92}**	1.18	0.05	**1.8×10^{-135}**
Cortex colour	S2_6566608	C	T	0.01	0.81	0.04	**5.0×10^{-83}**	0.89	0.04	**2.0×10^{-113}**

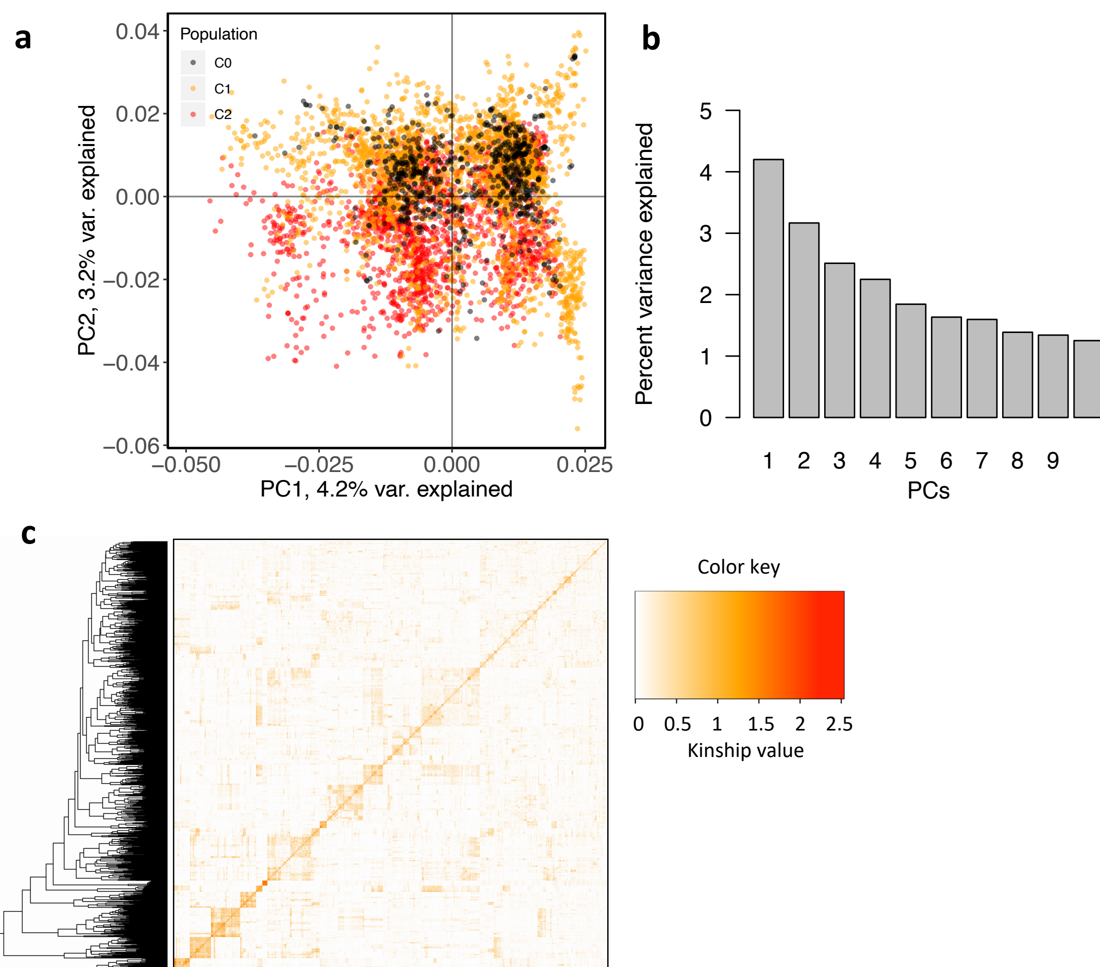
A1 = Reference allele (the coded effect allele), A2 = Alternate allele, MAF = Frequency of the reference allele, B = SNP effect, SE = standard error of SNP effect, P = marker-trait association p-value. Bold fonts represent markers that are significant at the Bonferroni threshold of 0.05/101,521 = 4.93×10^{-7}

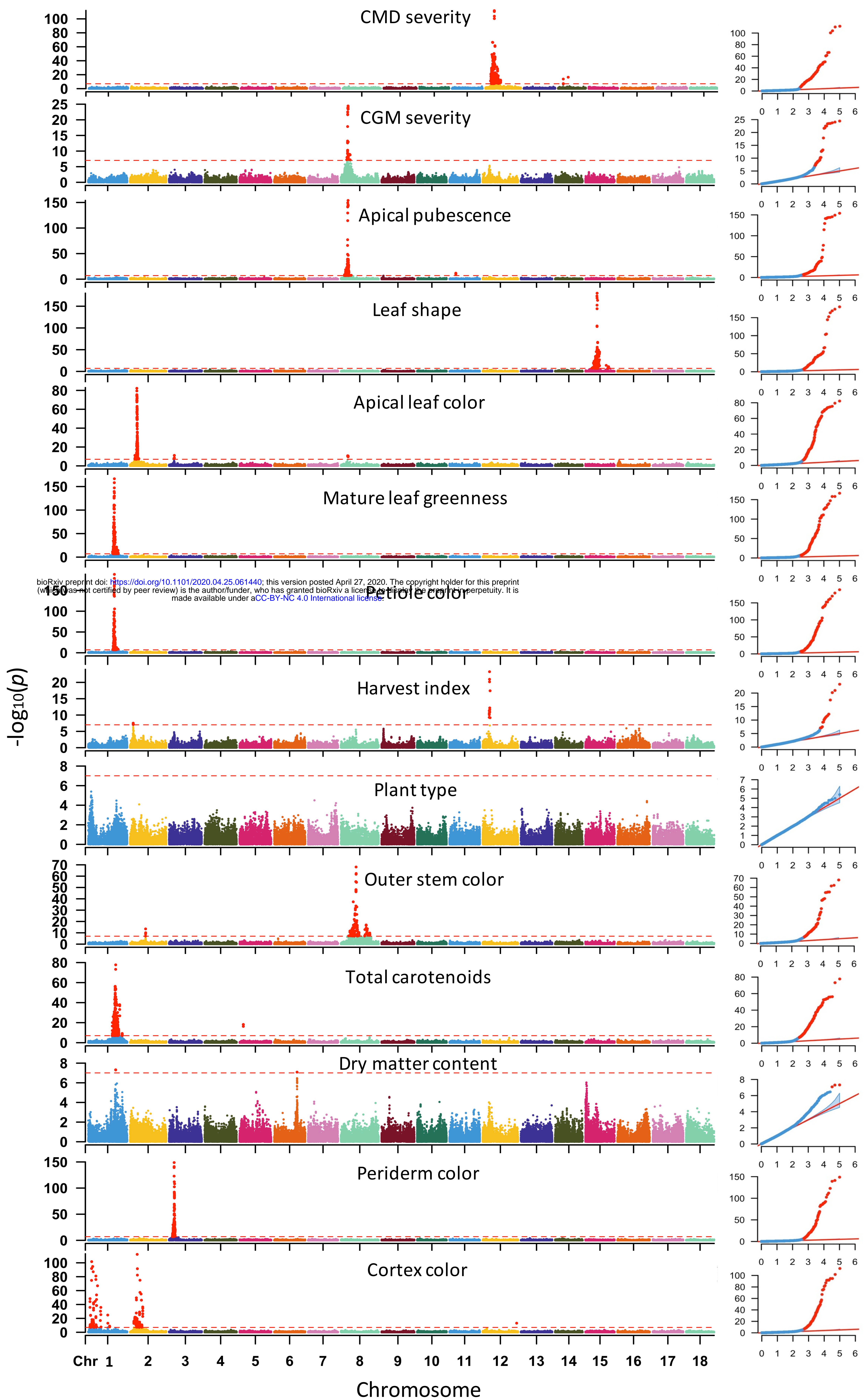
* SNP effect, standard error and p-values obtained from GCTA MLMe model.

	Mature_leaf_greeness	CGM_severity	TC_chart	Outer_stem_color	Root_cortex_color	CMD_severity	Petiole_color	Apical_pubescence	Dry_matter_content	Leaf_shape	Root_periderm_color	Plant_type	Harvest_index
Apical_leaf_color	0.1	0.06	0.03	-0.01	0.03	-0.13	0.01	-0.08	0.03	0.02	0.08	0.03	-0.02
Mature_leaf_greeness		0.06	0.17	0.08	-0.01	0	-0.65	-0.05	-0.04	-0.03	0	0.07	0.02
CGM_severity			0.05	0.07	0.04	0.04	-0.07	-0.27	-0.03	0.04	-0.04	-0.1	0.06
TC_chart				0.12	0.27	-0.07	-0.22	0.04	-0.31	-0.01	-0.12	-0.18	-0.05
Outer_stem_color					0.04	-0.01	-0.05	-0.02	-0.05	-0.05	-0.07	0.02	-0.11
Root_cortex_color						-0.03	0.01	-0.06	-0.04	0.01	0.08	-0.02	-0.09
CMD_severity							0.02	-0.12	-0.08	-0.07	0.02	0.08	-0.14
Petiole_color								0.06	0.06	0	-0.01	0.02	-0.05
Apical_pubescence									0.05	0	-0.04	0.13	0.01
Dry_matter_content										0.01	0.13	0.07	0.01
Leaf_shape											0.01	0.05	0.05
Root_periderm_color												0.08	-0.07
Plant_type													0.05

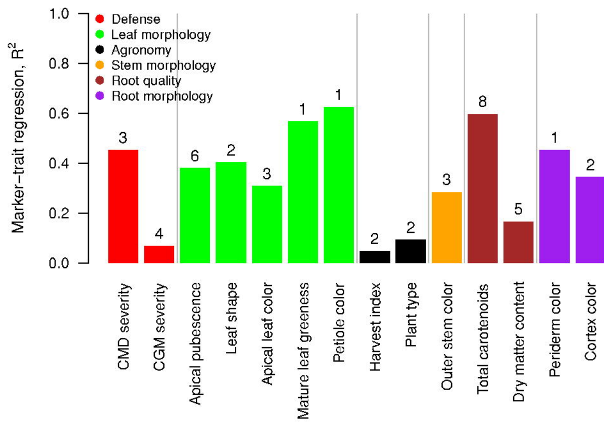

a
0Mb
8Mb
16Mb
23Mb
31Mb
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
0
1
21
41
61
81
101
121
141
161
>161

b
Count
0
2000
4000
6000
8000
0.0
0.1
0.2
0.3
0.4
0.5
Minor allele frequency

c
LD (r^2)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
100
200
300
400
500
600
700
800
900
Distance (kilobases)

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.





bioRxiv preprint doi: https://doi.org/10.1101/2020.04.25.061440; this version posted April 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



Marker−trait regression, R^2
1.0
0.8
0.6
0.4
0.2
0.0
Defense
Leaf morphology
Agronomy
Stem morphology
Root quality
Root morphology
3
4
6
2
3
1
1
2
2
3
8
5
1
2
CMD severity
CGM severity
Apical pubescence
Leaf shape
Apical leaf color
Mature leaf greeness
Petiole color
Harvest index
Plant type
Outer stem color
Total carotenoids
Dry matter content
Periderm color
Cortex color