Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Title: Regional Heritability Mapping provides insights into Dry matter (DM) Content in African white and yellow cassava populations.

Authors information (First names, surnames):

Uche Godfrey Okeke (First author),

Section of Plant breeding and Genetics, School of Integrative Plant Sciences, College of Agriculture and Life sciences, Cornell University, 14853, Ithaca, New York, USA.

Email: **urchgene@gmail.com**

Deniz Akdemir*,

Section of Plant breeding and Genetics, School of Integrative Plant Sciences, College of Agriculture and Life sciences, Cornell University, 14853, Ithaca, New York, USA.

Email: deniz.akdemir.work@gmail.com

Ismail Rabbi,

International Institute of Tropical Agriculture (IITA), PMB 5320, Oyo Road, Ibadan, Nigeria.

Email: i.rabbi@cgiar.org

Peter Kulakow,

International Institute of Tropical Agriculture (IITA), PMB 5320, Oyo Road, Ibadan, Nigeria.

Email: p.kulakow@cgiar.org

Jean-Luc Jannink, (corresponding author)

USDA-ARS, Robert W. Holley Centre for Agriculture and health, Tower Road, Ithaca, New York, 14853-2901.

Section of Plant breeding and Genetics, School of Integrative Plant Sciences, College of Agriculture and Life sciences, Cornell University, 14853, Ithaca, New York, USA.

Email: **jeanluc.jannink@ars.usda.gov**

***** Current affiliation: Statgen Consulting, Ithaca, NY 14850, USA.

Core ideas:

- Regional heritability mapping (RHM) is effective for understanding the genetic architecture of complex traits in cassava.
- Prediction accuracies can reflect the impact of genomic segments on cassava dry matter (DM) content.
- Serine-threonine protein kinases (SnRKs) are candidates positionally associated with cassava DM.
- Prediction accuracy of SnRKs for cassava DM was 50% of the total accuracy from genome-wide SNPs.

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Abstract:

The HarvestPlus program for cassava (*Manihot esculenta* Crantz) fortifies cassava with beta-carotene by breeding for carotene-rich tubers (yellow cassava). However, a negative correlation between yellowness and dry matter (DM) content has been identified. Here, we investigated the genetic control of DM in white and yellow cassava subpopulations. We used regional heritability mapping (RHM) to associate DM to genomic segments in both subpopulations. Significant segments were subjected to candidate gene analysis and we attempted to validate candidates using prediction accuracies. The RHM procedure was validated using a simulation approach. The RHM revealed significant hits for white cassava on chromosomes 1, 4, 5, 10, 17 and 18 while hits for the yellow were on chromosome 1. Candidate gene analysis revealed genes in the carbohydrate biosynthesis pathway including the plant serine-threonine protein kinases (SnRKs), UDP-glycosyltransferases, UDP-sugar transporters, invertases, pectinases, and regulons. Validation using 1252 unique identifiers from the SnRK gene family genome-wide recovered 50% of the predictive accuracy of whole genome SNPs for DM while validation using 53 likely (extracted from literature) genes from significant segments recovered 32%. Genes including an acid invertase, a neutral/alkaline invertase and a glucose-6-phosphate isomerase were validated based on an *a priori* list for the cassava starch pathway and also a fructose-biphosphate aldolase from the calvin cycle pathway. The power of the RHM procedure was estimated at 47 percent when the causal QTL generated 10% of the phenotypic variance with sample size of 451. Cassava DM genetics is complex. RHM may be useful for complex traits.

List of Abbreviations:

Dry matter content (DM)

Fresh root yield (FYLD)

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Linkage disequilibrium (LD)

Quantitative trait loci (QTL)

Regional heritability mapping (RHM)

Serine-threonine protein kinases (SnRKs)

Single nucleotide polymorphisms (SNPs)

Genotype-by-sequencing (GBS)

Genome-wide association analysis (GWAS)

Background:

Cassava currently ranks as the sixth world staple crop consumed by more than 500 million people in Africa, Asia and South America (El-Sharkawy, 2003). It was originally a perennial shrub but is cultivated now as an annual for its starchy root (El-Sharkawy, 2003). It is an outbreeding species and considered to be an amphidiploid or sequential allopolyploid (El-Sharkawy, 2003). The crop is clonally propagated by mature woody stem cuttings called stakes, which are 15-30 cm long and planted mostly inclined on ridged soils (Keating et al., 1988). Botanical seeds are used mainly in breeding programs with up to three seeds produced per pod (Iglesias et al., 1994, Iglesias and Hershey, 1991). Storage roots are generally harvested 7–24 months after planting (El-Sharkawy, 2003). Dry matter (DM) is the major product from cassava roots apart from moisture and traces of water-soluble vitamins and pigments (Holleman and Aten, 1956; Barrios and Bressani, 1967; Lim, 1968). On average, cassava DM is made up of about 90% carbohydrates (mainly starch), 2% protein, 1% fat, 3% minerals and ash and 4% fiber (Holleman and Aten, 1956; Barrios and Bressani, 1967; Lim, 1968). This starch deposit makes cassava attractive for the food industry and other industries that rely heavily on starch as their primary raw material (Lim, 1968). The value of cassava derives from a combination of fresh root yield and the percentage DM that can be extracted from fresh roots, referred to as dry yield. Fresh cassava roots with high DM content are also preferred by local farmers and processors (Kawano et al., 1987; Safo-Kantanka and Owusu-Nipah, 1992; Enidiok et al., 2008) who transform cassava roots into valuable staples consumed by many in developing countries. With 263 million metric tons produced in 2012 (FAOSTAT Database, 2013), cassava has become an

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

indispensable staple in the world and improvement of cassava for high dry yield is needed. This improvement should also endeavor to increase micronutrient content, as it is much needed in the cassava consuming regions of the world. Biofortification is a successful genetic improvement technique for increasing micronutrient content in staple crops (Meenakshi et al., 2010; Bouis et al., 2011) and represents a promising approach for solving the problem of micronutrient malnutrition around the world (Meenakshi et al., 2007; Meenakshi et al., 2010; Pfeiffer and McClafferty, 2007).

The target of biofortification is to increase the content of essential micronutrients such as Iron, Zinc, and Vitamin A (Meenakshi et al., 2007; Meenakshi et al., 2010; Pfeiffer and McClafferty, 2007), hence improving the health of millions of people who depend on these staples for daily nutrition. The biofortification process is facilitated by plant breeding (Meenakshi et al., 2010; Bouis et al., 2011). Since the early 2000s, the HarvestPlus initiative (Meenakshi et al., 2007; Pfeiffer and McClafferty, 2007) has been tasked with biofortification of staple crops including cassava, sweet potato, maize, rice and wheat. Biofortification of cassava is geared towards breeding varieties containing increased levels of provitamin A, or beta-carotene, a precursor for vitamin A. The so-called ‘yellow cassava’ (Liu et al., 2010; Plus, 2009; Aniedu and Omodamiro, 2012; La Frano et al., 2013) is designed to address public health issues including child mortality, impaired vision and night blindness, reduced immunity to diseases and other consequences of vitamin A deficiency (Liu et al., 2010; Plus, 2009).

Breeding for required levels of provitamin A necessitates the accumulation of beta-carotene in cassava roots (Aniedu and Omodamiro, 2012; La Frano et al., 2013). Many breeding programs use yellow flesh color as a proxy for measuring beta-carotene amount in cassava despite the fact that yellowness is more of an indication of total carotenoids in the root (Chávez et al., 2005; Ssemakula et al., 2007; Akinwale et al., 2010). This protocol is used to visually pre-select lines containing beta-carotenoids prior to quantification of different carotenoid levels using HPLC protocols (Kimura et al., 2007; Adewusi and Bradbury, 1993). Breeding for farmer preferred bio-fortified cassava involves the development of high yielding clones with high DM and high beta-carotene accumulation in a single clone or variety (Ceballos et al., 2004; Raji et al., 2007). Incorporating all these characteristics in a single variety of cassava makes for a challenging breeding task. Some studies have shown that there is a negative genetic correlation between DM and yellow root flesh color in cassava making this breeding task even more challenging since the target is towards full adoption of pro-vitamin A varieties by local farmers and processors (Akinwale et al., 2010; Vimala et al., 2008). It is therefore useful to understand the genetic control of DM content and beta-carotene accumulation in cassava to facilitate the breeding of farmer-preferred varieties.

Regional heritability mapping (RHM) is a relatively new procedure for identifying loci affecting quantitative traits (Nagamine et al., 2012; Riggio and Pong-Wong, 2014; Riggio et al., 2013; Shirali et al., 2015). Unlike single marker GWAS methods which the lack power to detect rare genetic variants (Bodmer and Tomlinson, 2010; Gibson, 2012; Wood et al., 2014), RHM can capture both rare and common genetic variants giving it

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

more power to identify loci that cannot be detected by standard GWAS (Nagamine et al., 2012; Riggio and Pong-Wong, 2014; Riggio et al., 2013). The RHM has been shown to detect both common and rare genetic variants implicated in disease traits in human genomics (Shirali et al., 2015; Uemoto et al., 2013; Zeng et al., 2016) and recently in tree genomics (Resende et al., 2017). RHM is a suitable method for capturing the effect of a genomic block or segment since it can identify genomic segment-trait associations for regions spanning multiple loci (Nagamine et al., 2012; Riggio and Pong-Wong, 2014; Riggio et al., 2013; Caballero et al., 2015). A multi-marker mapping approach like the RHM may identify both common and rare variants involved in the expression of DM in white and yellow subpopulations of African cassava. To the best of our knowledge, this is the first attempt to use the RHM procedure in an annual crop.

The objectives of this study were:

1. To understand the genetic basis of DM in white and yellow root African cassava populations.
2. To determine the power of the RHM procedure to detect genomic segments carrying QTL using the hide-a-causal-SNP procedure.

Materials and Methods:

Cassava phenotypic data for discovery:

We used phenotypic data collected from the Genetic Gain (GG) population trials conducted by the cassava breeding program at the Institute of Tropical Agriculture (IITA), Ibadan, Nigeria for our analysis. The GG population (713 clones) is an elite population bred from the 1970s to 2007 by the cassava breeding program at the IITA (Maziya-Dixon et al., 2007; Okechukwu and Dixon, 2008; Ly et al., 2013). Most GG clones are of African origin with very good performance such that they were advanced to advanced to multi-environment uniform yield trials. For this study, we used clonal evaluation trials (CETs) of the GG population planted in an augmented design. The CET uses an unreplicated incomplete block design consisting of a layout of between 18 to 30 blocks with 22 accessions and two checks in each block. Accession plots were a single row (1m x 1m spacing) of five-plant stands without borders. All checks were included in the analysis. A few trials were replicated twice. These trials were conducted in three locations in Nigeria: Ibadan (7.40° N, 3.90° E), Mokwa (9.3° N, 5.0° E), and Ubiaja (6.66° N, 6.38° E) between 2013 and 2015. Three core agronomic traits were measured for these trials including fresh weight of harvested roots expressed in tons per hectares (T/ha) (FYLD), percentage dry matter (DM) of storage roots, which measures root dry weight as the percentage of the root fresh weight, and pulp color (PLPCOL) a binary trait rated on a scale from 1 (white flesh to light cream root) to 2 (deep cream to yellow flesh root). The DM trait was measured using the oven method: 100g grated root sample (with thorough mixing of 10-15 randomly selected roots from a plot) were collected per accession and oven dried. DM content was then measured as residual

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

weight after oven drying. We further divided the GG population (713 clones) into two subpopulations of white (451 clones) and yellow (262 clones) cassava using the PLPCOL trait where clones with a score of 1 for this trait were grouped into the white population and those with score 2 into the yellow population.

Cassava phenotypic data for validation:

To validate results from the RHM analysis, we used data from a population called the GS-C, which consisted of progenies of clones from the GG population described above. Phenotypes from the GS-C1 were obtained from clonal evaluation trials (CETs) of 1,651 clones split into trials at three locations: Ibadan, Mokwa and Ikenne (6°52′N 3°43′E). These trials were planted using an augmented design consisting of between 20 to 30 blocks with 22-24 clones and two checks in each block. Plots were a single row of five-plant stands (1m x 1m spacing) without borders and without replication and trials were planted during 2014 and 2015. Cassava trait measurements for this population were as described earlier, except that no strict distinction between yellow and white flesh color was used because the GS-C1 were majorly white and cream clones; thus we performed validation analysis using all clones.

Cassava genotype data:

DNA was extracted using DNeasy Plant Mini Kits (Qiagen) from 713 clones from the 2013 Genetic Gain trial at IITA and was quantified using PicoGreen. Genotyping-by-sequencing (GBS) was used for genotyping (Elshire et al., 2011) these clones. Six 95-plex and one 75-plex *Ape*KI libraries were constructed and sequenced on Illumina

HiSeq, one lane per library. Single nucleotide polymorphisms (SNPs) were called from the sequence data using the TASSEL pipeline version 4.0 (Glaubitz et al., 2012), using an alignment to the *M. esculenta* version-6 reference genome (Goodstein et al., 2012). The marker data was converted to dosage format (0, 1, 2) and missing genotypic data were imputed using the Beagle software (Ayres et al., 2011). The final data set consisted of 177,201 SNPs scored in 713 clones. Members of the GS-C1 used in the validation analysis were genotyped in 2014 as described above. SNPs from both populations were called together using the TASSEL pipeline (Glaubitz et al., 2012) and missing genotypes also imputed using Beagle (Ayres et al., 2011) yielding the same number of SNPs as above.

Data analysis:

Genome-wide Regional Heritability mapping (RHM):

RHM was carried out using the following procedure (Nagamine et al., 2012; Riggio and Pong-Wong, 2014; Riggio et al., 2013):

a. Chromosomes were divided into 100 SNP segments in sliding windows with 50 SNPs overlapping between adjacent windows.
b. A multikernel univariate mixed model was used to partition the genomic additive variation due to trait of interest into components of the target genomic segment and the whole genome SNP markers as follows:

$$y = X\beta + Zu_1 + Zu_2 + e \qquad (1)$$

$$u_1 \sim N(0, \sigma^2 u_1 K u_1); \qquad u_2 \sim N(0, \sigma_{u_2}{}^2 K u_2); \quad e \sim N(0, \sigma_e{}^2 I_{n\times n})$$

$$var(y) = V = Z(\sigma^2 u_1 K u_1)Z^T + Z(\sigma^2 u_2 K u_2)Z^T + \sigma^2_e I$$

$$\hat{u}_1 = (\sigma^2{}_{u_1} K_{u_1})Z^T V^{-1}(y - X\hat{\beta}) ; \quad \hat{u}_2 = (\sigma^2{}_{u_2} K_{u_2})Z^T V^{-1}(y - X\hat{\beta})$$

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Where *y* is a response variable (DM), *X* is a known incidence matrix for fixed effects *β* (including grand mean and a nested effect of Rep within Trial within Year within Location), *Z* is a known incidence matrix for clonal additive genomic effects u_1 for the target genomic segment and u_2 for the whole genome SNPs. K_{u1} and K_{u2} are the genomic relationship matrices calculated from the SNPs using the procedure of VanRaden (2008) as:

$$G = \frac{MM^T}{2\Sigma p(1-p)}$$

where *G* is the genomic relationship matrix, *M* is a centered marker matrix coded as -1,0,1 and *p* is the major allele frequency vector. Other components of the model include the genomic variance for the target genomic segment ${\sigma_{u_1}}^2$ and the total genomic variance for the whole genome ${\sigma_{u_2}}^2$, ${\sigma_e}^2$ is the genomic error variance and e are the residuals from the model. Model (1) was fit using the R EMMREML package (Akdemir and Okeke, 2014). Note that the K_{u2} genomic relationship matrix serves to statistically control for population structure effects as the kinship matrix does in standard GWAS.

c. Following model fit from step (b) above, genomic heritability for each target genomic segment was computed as follows:

$$\boldsymbol{h^2} = \frac{{\boldsymbol{\sigma_{u_1}}}^2}{({\boldsymbol{\sigma_{u_1}}}^2 + {\boldsymbol{\sigma_{u_2}}}^2 + {\boldsymbol{\sigma_e}}^2)}$$

where $\boldsymbol{h^2}$ is genomic heritability for a target genomic segment and variance components are described above.

d. A likelihood ratio test (LRT) was used to test the significance of target genomic segments with the alternative model as model (1) and the null model as model

(1) without the target genomic kernel component ie $y = X\beta + Zu_2 + e$. This model was also fit using the EMMREML package (Akdemir and Okeke, 2014). P-values were obtained using the *pchisq* function in R (R Core Development Team, 2016).

e. Local FDR (LFDR) was estimated using the R qvalue package (Sorey and Tibshirani, 2003; Storey et al., 2015).
f. Genomic segment LFDRs were then plotted across the genome in a Manhattan plot with a cutoff of 0.05 used to assess significance.

We carried out the RHM procedure separately for the white and yellow cassava subpopulations of GG. No defined population structure was found on in the GG population in a previous GWAS study (Wolfe et al., 2015). Therefore, the genomic relationship matrix from the whole genome SNPs in the RHM was sufficient to account for structure in this analysis (in fact we refer to this more as background effect).

Candidate gene analysis:

We identified candidate genes from the significant hits of the RHM analysis based on annotations for the v6 *M. esculenta* genome on phytozome (Goodstein et al., 2012). We used plant physiology information to narrow down the list of genes associated with carbohydrate biosynthesis including genes functional in starch and sugar biosynthesis, cell wall loosening and degradation, and root sink and plant growth pathways. We carried out validation tests on selected candidates based on prediction accuracies on the GS-C1 population as described below.

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Validation models and procedures:

We conducted validation analyses for the significant hits of the RHM analysis and for the RHM procedure itself. Validation here was geared towards understanding the prediction accuracies obtained from genes and gene families on RHM significant segments. Validation proceeded as follows:

(a) Validation using SnRK genes (a candidate gene family):

To obtain genotypic data for this analysis, we searched the Phytozome *M. esculenta* v6.1 web portal (Goodstein et al., 2012) using the keyword *serine threonine kinases* to recover all its instances in the cassava genome, resulting in 2,408 hits. We filtered the resulting list to remove all hits not containing gene ontology or Eukaryotic Orthologous Groups function definitions for the keyword *serine threonine kinase*. We then manually added genes containing known serine threonine kinases that did not contain a function definition, for example the *SNF1* gene (a list of these genes is provided in the Supplementary Table). We extracted all markers within 2.5 kb of the start and end of each gene model using the Bedtools intersect function (Quinlan and Hall, 2010) resulting in 7,203 unique SNPs. We refer to these SNPs as candidate SNPs below. For validation of these candidate SNPs on the GS-C1 data we fit the following model:

$$y = X\beta + Zs + Zg + e \qquad (2)$$

$$s \sim N(0, \sigma^2_s K_s); \qquad g \sim N(0, \sigma^2_g K_g) \qquad e \sim N(0, \sigma_e{}^2 I_{n\times n})$$

$$var(y) = V = Z(\sigma^2_s K_s)Z^T + Z(\sigma^2_g K_g)Z^T + \sigma^2_e I$$

$$\hat{s} = (\sigma^2{}_s K_s)Z^T V^{-1}(y - X\hat{\beta}); \qquad \hat{g} = (\sigma^2{}_g K_g)Z^T V^{-1}(y - X\hat{\beta})$$

Where y is a vector of the raw phenotypic values for DM, X is the known incidence

matrix for fixed effects β (including grand mean and a nested effect of Trial within Year within Location), Z is known incidence matrix for clonal additive candidate genomic effects s and whole genomic effects g. For K_s, and K_g we used the candidate SNPs and the remaining SNPs from the whole genome excluding the candidate SNPs, respectively, to generate genomic relationship matrices for the 1,651 clones of the GS-C1 population as above. A third kinship matrix, K_{rand}, was generated as a control from 7,203 SNPs anchored to 2000 randomly selected genes from the cassava genome and used in Model (2) in place of $K_{s,}$ while we calculated K_g using SNPs from the whole genome excluding those in K_{rand}. Other components of the model include the SnRKs candidate genetic variance σ_s^2 and the genetic variance from other parts of the genome σ_g^2, σ_e^2 is the error variance and e is the residuals from the model. Model (2) was fit using the EMMREML. To assess prediction accuracies, we fit another model as follows:

$$y = X\beta + Zu + e \qquad (3)$$

$$u \sim N(0, \sigma^2_u I);\ e \sim N(0, \sigma_e^2 I_{n\times n})$$

$$var(y) = V = Z(\sigma^2_u K)Z^T + \sigma^2_e I; \qquad \hat{u} = (\sigma^2_u K)Z^T V^{-1}(y - X\hat{\beta})$$

where most components of Model (3) remain same as in Model (2) apart from the genetic effect u having an identity matrix I as its covariance matrix signifying that the 1,651 GS-C1 validation clones are unrelated. Model (3) was also fit using R EMMREML. Model (2) was fit using a 5-fold a cross validation (CV) scheme with 10 repeats and prediction accuracies were obtained for this CV scheme by a correlation of $\hat{s}$ of each clone from Model (2) to its $\hat{u}$ value from Model (3).

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

(b) Validation using 53 candidate genes extracted from plant physiology literature and 53 randomly selected genes from the RHM significant regions:

We performed a second procedure to validate the 53 candidate genes identified in significant hit regions in the RHM analysis based on plant physiology literature (Table 1). Using the cassava genome unique gene identifiers from Phytozome (Goodstein et al., 2012), we extracted all markers within 2.5Kb flanking the start and end of each gene as before, resulting in 400 unique SNPs. We refer to these SNPs as 'likely candidate SNPs'. We also picked 53 single copy genes at random from within the RHM significant regions and anchored them to 395 SNPs as controls for the likely candidate SNPs. We term these the 'unlikely candidate SNPs'. To validate these, we also fit the GBLUP Model (2) with these modifications: (1) for K_s we used K_{53} which was a genomic relationship matrix calculated from the 400 likely candidate SNPs for the 1,651 clones of the GS-C1 population (as above), (2) we calculated K_g using SNPs from the whole genome excluding these likely candidate SNPs, (3) K_{rand} was also calculated as above (as a control) from 402 SNPs anchored to 53 randomly selected genes from the cassava genome (with 7.5 kb flanking the start and end of these genes), (4) $K_{unlikely}$ was calculated from the 395 unlikely candidate SNPs. These were also used in place of K_s in Model (2) with their appropriate K_g calculated as other SNPs in the genome excluding those in K_{rand} and $K_{unlikely}$. Other components of the model were as described for Model (2) and prediction accuracies were obtained in the same way. To assess the prediction accuracy of the whole genome SNPs, we also fit a model analogous to Model (3) with covariance of u coming from a genomic relationship matrix with whole genome SNPs. We term this the predictive accuracy of the whole genome SNPs.

(c) Validation using all genes within 1Mb of the RHM significant list and an *a priori* list of starch genes in cassava:

We performed another validation procedure to provide a validation for all the genes identified in the significant hit regions in the RHM analysis, including those shown in Table 1 and those not shown because they were not selected on the basis of information from literature. Using the cassava genome unique gene identifiers from Phytozome (Goodstein et al., 2012), we extracted all SNPs within a 1 Mb region centered on each of these candidates using Bedtools resulting in 2,297 SNPs from 650 unique genes. We refer to these SNPs as all RHM region SNPs (RHM-regions). In addition we extracted SNPs anchored to 123 unique genes in the cassava starch pathway compiled by Saithong et al. (2013), resulting in 419 SNPs. We refer to these SNPs as cassava starch SNPs. To validate these SNPs, we fit Model (2) using genomic relationship matrices calculated as above from RHM-region and cassava starch SNPs, in place of K_s with their appropriate K_g calculated from remaining SNPs. We also picked 650 single copy genes at random excluding the RHM significant regions and anchored them to approximately 2300 SNPs as controls for the RHM-region and cassava starch SNPs. We refer to these as Random-650 SNPs. We calculated $K_{random-650}$ using these SNPs and an appropriate K_g. These kernels were also fit in Model (2) as K_s and K_g respectively. In addition to prediction accuracies from these candidates, we validated genes in the RHM-regions by searching for them in two *a priori* lists compiled by Saithong et al. (2013) including one for the cassava starch pathway and another for the Calvin Cycle pathway. RHM-region genes that made this list were considered

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

validated.

(d) Assessing the RHM power via the hide-a-causal-SNP procedure:

To validate the RHM procedure, we performed an analysis similar to the classical hide-a-causal-SNP approach as follows:

a. Chromosomes were divided into 100 SNP segments in sliding windows with 50 SNPs overlapping between adjacent segments.
b. Five (5) adjacent segments were randomly selected on each chromosome.
c. On the third segment, effects were added to a random SNP to inflate the phenotypic variance of the DM trait by 10%.
d. Genomic relationship matrices were made for these segments but for segment 3, the random pseudo-causal SNP was excluded when calculating the genomic relationship matrix.
e. Subsequently, steps (b) to (d) of the RHM procedure above were carried out, resulting in P-values for these five adjacent segments. Steps (a) to (e) were repeated twelve times, resulting in 216 tests.
f. We then calculated the P-value from the RHM analysis on our data that corresponded to the LFDR threshold of 0.05 and used this as significance threshold.
g. The power of the RHM analysis was then calculated as the number of times any of the five segment P-values were significant given the significance threshold from (f) above.
h. To make a decision on the bounds set for extracting adjacent candidate genes

from the *M. esculenta* genome for a significant segment in the RHM analysis, the number of times either the 1st or 5th segment P-values were significant conditional on the 3rd segment having a higher P-value were also calculated. This reflected how far away adjacent segments captured causal variants.

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Results:

RHM for DM in white and yellow cassava populations:

The genomic heritabilities for DM in white and yellow cassava based on whole genome SNPs were 0.57 and 0.48 respectively. These heritabilities are somewhat higher than those found by Ly et al. (2013), presumably because they worked with more locations and years and thus experienced higher genotype-by-environment interaction. We observed different genetic control patterns for DM in the white and yellow cassava subpopulations as shown by Manhattan plots from the RHM analysis (Figure 1). Significant genomic segments for the white cassava DM were observed on chromosomes 1, 4, 5, 10, 17 and 18 while for the yellow cassava a significant segment was only observed on chromosome 1 (Figure 1). Due to the difference between the sample sizes of both subpopulations, it is unclear if the DM genetic control patterns between these subpopulations were different. A non-significant but strong signal was also observed on chromosome 9 of both cassava subpopulations.

Candidate gene analysis:

Using information from the estimates of the mean LD between genomic segments per chromosome (Figure 2A), the distribution of the length of genomic segments in our analysis (Figure 2B) and information on the number of times adjacent segments captured causal variations in the simulation analysis; we set the bound for the region where candidate genes were sought to 1.0 Mb (500Kb flanking each hit), representing from two to three genomic segments adjacent to the top hit genomic segment.

Candidates for the white and yellow cassava subpopulations:

For the top RHM hits in both cassava gene pools, we identified possible candidate genes and transcriptional regulators adjacent to these hits based on their involvement in the carbohydrate biosynthesis pathway including members of the serine/threonine kinase family (SnRKs), members of the UDP-glycosyltransferase family (including starch and sucrose synthases), and UDP-sugar transporters, specific plant transcriptional factors including members of the beta helix-loop-helix (bHLH) family and mini zinc fingers, and other genes involved in cell wall processes, root storage and development including pectinases and beta vacuolar processing enzymes. We show a list of these genes in Table 1. An additional candidate gene, phosphofructokinase, was associated with the non-significant peak on chromosome 9 which was more pronounced in the yellow cassava germplasm.

Validation results for SnRKs:

The predictive accuracy of the whole genome SNPs was 0.54 (0.03). Using the set of candidate SnRK SNPs, prediction accuracies from the CV using Model (2) were 0.26 (0.04) and 0.12 (0.06) for the candidate and random SNPs, respectively, with standard deviation of the cross validation repeat cycles shown in parentheses. The predictive ability of the genome-wide SnRK candidates (7,203 SNPs) had approximately 50 percent of the total prediction accuracy from our set of genome-wide SNPs (177,201) for the GS-C1 population.

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Validation using 53 likely candidate genes extracted from plant physiology literature and 53 unlikely candidate genes from the RHM significant regions:

Using the likely candidate SNPs from the genes identified for all the top hit genomic segments genome-wide (shown in Table 1), prediction accuracies from the CV using a modified Model (2) were 0.17 (0.03), those for the 53 unlikely genes randomly selected from the top hit genomic segments genome-wide were 0.14 (0.02) and those for the SNPs from random 53 genes from the cassava genome were 0.06 (0.08) with standard deviation of the cross validation repeat cycles in parentheses.

Validation using all genes within 1Mb of the RHM significant list and an *a priori* list of starch genes in cassava:

Using the RHM-region, cassava starch and Random-650 SNPs, the prediction accuracies from the CV using a modified Model (2) were 0.17 (0.04), 0.18 (0.03) and 0.03 (0.01) respectively. Based on two *a priori* lists compiled by Saithong et al. (2013) including one for the cassava starch pathway and another for the Calvin cycle pathway, we found three RHM-region genes on the cassava starch pathway list including an acid invertase (*Manes.01G076500*), a glucose-6-phosphate isomerase (*Manes.18G060600*) and a neutral or alkaline invertase (*Manes.04G006900*). However, from the Calvin Cycle pathway list we found one RHM-region gene, namely fructose-biphosphate aldolase (*Manes.04G007900*). These genes are known to play key roles in starch biosynthesis and storage (Junker, 2004; Ap Rees, 1992; Appeldoorn et al., 1997; Renz et al., 1993). To assess if these genes were significantly enriched in RHM regions, we performed a simple calculation by multiplying the 650 genes in the RHM region with 123

genes in the cassava starch pathway (Saithong et al., 2013) and divided them by the total number of genes in the cassava genome (33,030). The result was 2.4, which is the expectation of a Poisson process of obtaining the genes in the cassava starch pathway. However we calculated the probability of drawing 3 cassava starch pathway genes from the genome at random resulting in p = 0.22 indicating no significant enrichment.

Assessing the RHM power via the hide-a-causal-SNP procedure:

We calculated the statistical power of the RHM procedure to detect simulated causal effects from 216 analyses as the number of times any of the five segment P-values were significant. The P-value from the RHM analysis on our data that corresponded to the LFDR threshold of 0.05 was 0.00024, which became our significance threshold for this analysis. We found that 102 tests were significant out of a total of 216 representing a 47 percent statistical power to detect the simulated causal region. To set the bounds for how far in the genome to cover when extracting candidate genes from an RHM significant segment, we also calculated the number of times P-values from the 1st or 5th genomic segments were significant conditional on the 3rd segment's P-value being higher. With a total of 216 analysis, 27 cases had significant P-values on segment 3 and 15 cases had significant P-values from segments 1 or 5 when the P-values from segment 3 were higher. This represents 15 percent coverage farther away from the causal segment. With this information we chose an adjacent span of 500,000 kb pairs flanking an RHM significant segment as the bounds for extracting adjacent candidate genes.

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Discussion:

The RHM results in the high DM and white cassava populations clearly demonstrate the polygenic nature of the DM trait. DM is composed of carbohydrates (mostly starch), cell wall components and fiber, as well as other non-starchy polysaccharides. Thus, it is not surprising that this trait is complex and controlled by many genes. Also the RHM procedure in this study showed a 47% power for detection of association with a sample size of less than 500 given the polygenic nature of this trait.

SnRKs may be involved in regulation of cassava carbohydrate biosynthesis:

The serine/threonine protein kinase (SnRKs) gene family in plants is homologous to the sucrose non-fermenting 1 (SNF1) protein kinase family in yeast and the AMPK gene family in mammals. Its members have gained recognition as critical elements in transcriptional, metabolic and developmental regulation in plants (Halford et al., 2003; Halford et al., 1998; Polge and Thomas, 2007; Xue-Fei et al., 2012; Crozet et al., 2014; Jossier et al., 2009). The most studied member of this family is the SnRK1 (Halford et al., 1998; Polge and Thomas, 2007). SnRKs play a vital role as global regulators of carbon metabolism and mediate cross talk between metabolic and other plant signaling pathways (Halford et al., 1998; Polge and Thomas, 2007; Xue-Fei et al., 2012). SnRK1 was shown to play a key role in seed filling and maturation and in embryo development in peas (Radchuk et al., 2010; Radchuk et al., 2006). In potato and wheat, SnRK1 phosphorylates and inactivates key enzymes in the sugar and starch biosynthesis pathway, affecting sucrose synthase, trehalose phosphate synthase and alpha amylase (Purcell et al., 1998; Laurie et al., 2003), and in potato, it stimulates the redox activation of ADP-glucose pyrophosphorylase (AGPase) in response to high sucrose levels

(Geigenberger, 2003; Tiessen et al, 2003). Antisense expression of SnRK1 resulted in a reduction in the expression of sucrose synthase in potato tubers (Purcell et al., 1998) and alpha amylase in cultured wheat embryos (Laurie et al., 2003). However, the overexpression of SnRK1 in potatoes resulted in a significant increase in starch accumulation in tubers and a decrease in glucose levels resulting from a dramatic increase in the activity and expression levels of sucrose synthase and AGPase (McKibbin et al., 2006). SnRK1 is activated by high cellular sucrose and/ or low glucose or a dark period (Rolland et al., 2002). The model of sugar and starch biosynthesis in potato from McKibbin et al. (2006) showed SnRK1 at the heart of these processes. Using RHM analysis in the white cassava population, we identified significant genomic segments containing some of the proteins or enzymes in the model given in this illustration (McKibbin et al., 2006) including SnRKs, UDP-Glycosyltransferases and UDP-sugar transporters, an ADP-type starch synthase 2 and a neutral invertase. Glycosyltransferases are a family of enzymes involved in carbohydrate biosynthesis of which sucrose and starch synthases are members (Momma and Fujimoto, 2012). Using the RHM procedure and candidate gene analysis, several of these known carbohydrate biosynthesis enzymes (Table 1, Figure 3) were putatively associated with the cassava DM trait.

Other possible candidates that are involved in sugar and starch biosynthesis in Cassava:

Other proteins located within significant genomic segments that are also involved in the carbohydrate biosynthesis pathway include invertase inhibitors which have been shown to form complexes with SnRKs and lead to reduced accumulation of reducing sugars

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

and increased accumulation of starch in potatoes (Lin et al., 2015), and BAK1, a brassinosteroid insensitive 1 (BR1) associated receptor-like kinase and a member of the somatic embryogenesis receptor-like kinase (SERKs) subfamily involved in regulation of root development (Du et al., 2012). BAK1/serk1 positively controls starch granule accumulation in Arabidopsis root tips (Du et al., 2012). Using a transgenic sweet potato overexpressing a DNA-binding one zinc finger (Dof) protein encoded by a *SRF1* gene (a member of the mini zinc finger family of plant specific transcription factors (Takatsuji, 1998; Takatsuji, 1999)), Tanaka et al. (2009) showed that transgenic roots had significantly higher storage root dry matter content, increased starch content per fresh weight of storage root and a drastic decrease in glucose and fructose levels (Tanaka et al., 2009). *SRF1* was shown to modulate carbohydrate metabolism in sweet potato storage roots via negative regulation of vacuolar invertase (Tanaka et al., 2009). Several enzymes, including pectinases, pectin esterases, cellulase synthase and galacturonosyltransferases (GAUT), found in the RHM significant regions in white and yellow cassava may be involved in plant cell wall loosening and degradation which may be linked to carbon partitioning in cassava. In fact GAUT, a member of the CAZy (Cantarel et al., 2009) GT8 family of glycosyltransferases, is involved in pectin and hemicellulose biosynthesis (Cantarel et al., 2009; Atmodjo et al., 2011; de Godoy et al., 2013). GAUT-silenced tomato fruits showed altered pectin composition and decreased starch accumulation (de Godoy et al., 2013). Cassava GAUTs may interfere with carbon metabolism, partitioning and allocation as seen in tomato (de Godoy et al., 2013). In their expression profile study using samples from different stages of cassava root development, Yang et al. (2011) found a significant up-regulation of these enzymes

involved in plant cell wall loosening and degradation. The beta helix-loop-helix (bHLH) family of transcription factors is a large family in plants involved in flavonoid, carotenoid pathway and anthocyanin pigmentation of tuber skin and flesh (from yellow to white and purple) in potato (De Jong et al., 2004; Zhang et al., 2009; Tai et al., 2013) and may interact with sucrose transporter to perform this function (Krügel et al., 2012). Phytochrome-interacting factors (PIFs) form a subfamily of bHLH transcription factors and PIF1 (a member of this subfamily) have been shown to directly regulate the expression of phytoene synthase (PSY) (Toledo-Ortiz et al., 2010), a major driver of carotenoid production in plants and the first and main rate-determining enzyme of the carotenoid pathway (Toledo-Ortiz et al., 2010; Maass et al., 2009). It is not clear how bHLH may link with sugar biosynthesis and transport or play a role in starch accumulation in yellow cassava clones, but this may translate to the frequently observed negative correlation between DM and yellow root flesh color in African cassava (Esuma et al., 2016; Akinwale et al., 2010). Interestingly, cassava breeders in Colombia have not found any negative correlation between carotenoids and DM in their germplasm and in fact have made gains in both traits using a rapid cycling recurrent selection scheme (Ceballos et al., 2013).

Some experimental studies that reflect possible roles of candidate genes in the cassava tuber:

Using the RHM analysis, we identified (Figure 3) a number of cassava genes in the heterotrophic plant cell starch/sucrose metabolism pathway (Junker, 2004). We describe a few steps in this pathway, concentrating mostly on where we have identified candidate genes (candidate genes are in braces henceforth with phytozome gene

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

identifiers). After sucrose is imported into the cytosol by a sucrose transporter (Manes.05G099000, Manes.18G054200), it is converted into hexose sugars via two paths involving the enzymes sucrose synthase (shown in the center of Fig. 3) and invertase (shown to the left in Fig. 3) (Manes.04G006900, Manes.01G076500) (Junker, 2004; Ap Rees, 1992; Appeldoorn et al., 1997; Renz et al., 1993).

Sucrose transport is much more pronounced in the sink tissues that switch to storage mode (Weschke et al., 2000; Weschke et al., 2003). A transgenic study using *sucrose transporter 4-RNAi* potato plants showed an increase in tuber yield and starch accumulation, and also induced early tuberization (Chincinska et al., 2008). It is worth noting that the cytosolic neutral invertase tends to play a larger role in sink organs than does the vacuolar acid invertase. Studies on maize null mutants of the cytosolic invertase (*Mn1*) had miniature seeds due to arrested endosperm development (Miller and Chourey, 1992), while overexpression of *Mn1* increased grain yield and starch content (Li et al., 2013). Similar studies in rice, tomato and cotton have also found consistent phenotypes with cytosolic neutral invertase (Wang et al., 2008; Zanor et al., 2009; Wang and Ruan, 2012). Other studies on vacuolar invertase inhibitors showed a significant reduction of cold-induced sweetening in potato tubers (via a reduction in sucrose accumulation in tubers) by restricting the activities of vacuolar acid invertase (McKenzie et al., 2013; Brummell et al., 2011). These studies suggest the importance of sucrose unloading to sink organs and hence vacuolar acid and cytosolic invertases are targets for post-translational regulation towards starch storage and dry matter accumulation (Tang et al., 2016).

The hexoses cleaved from sucrose are rapidly phosphorylated into hexose monophosphates by hexokinase and fructokinase (Junker, 2004; Ap Rees, 1992; Appeldoorn et al., 1997; Renz et al., 1993) and they proceed to starch biosynthesis or glycolytic pathways. As shown in the central pathway in Figure 3, the resulting hexose monophosphates (including glucose-1-phosphate, glucose-6-phosphate and fructose-6-phosphate) are interconverted by the enzymes phosphoglucose mutase and phosphoglucose isomerase (Manes.18G060600) (Junker, 2004). Phosphoglucose isomerase connects the Calvin Cycle pathway with the starch biosynthetic pathway in illuminated plant leaves (Bahaji et al., 2015). It also plays a key role in the glycolytic pathway and in the regeneration of glucose-6-phosphate in the oxidative pentose pathway in heterotrophic organs and non-illuminated plant leaves (Bahaji et al., 2015). It is strongly inhibited by light (Heuer et al., 1982) and by an intermediate Calvin Cycle molecule 3-phosphoglycerate (3PGA) (Dietz, 1985), which accumulates in the chloroplast during illumination and allosterically activates AGPase (Kleczkowski, 1999; Kleczkowski, 2000). The second phosphorylation step in the glycolytic pathway is the phosphorylation of fructose-6-phosphate to fructose-1,6-bisphosphate by phosphofructokinase (Manes.09G077800). Interestingly, transgenic studies overexpressing 6-phosphofructokinase in potato found no changes in the transgenic tuber phenotype compared to the controls but had an increased flux of cytosolic 3PGA that did not affect the amount of starch that accumulated in the tubers (Sweetlove et al., 2001; Burrell et al., 1994). It is noteworthy that our RHM results identified a signal on chromosome 9 in both yellow and white cassava that corresponds to the position of a phosphofructokinase in cassava.

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Fructose-bisphosphate aldolase (FDA), a candidate from the Calvin Cycle pathway (*Manes.04G007900*), is known to play a key role in carbohydrate biosynthesis. Changes in FDA activity have marked consequences for photosynthesis, carbon partitioning, growth, yield and improved uniformity of solids in potato and other plants (Haake et al., 1998; Barry et al., 2002). Transgenic plants (including potato, corn, rice, canola and other crops) that expressed the *E. coli* FDA gene in their chloroplasts had significantly higher root mass, leaf phenotypes with significantly higher starch accumulation, and lower leaf sucrose compared to control plants expressing the null vector (Barry et al., 2002).

Result implications for the breeding of high DM white cassava varieties or high DM, high beta carotene yellow cassava varieties

The RHM results presented in this study suggest that DM content is under complex genetic control, particularly in the white cassava population. A network of genes and transcriptional regulons that are at the heart of sugar and starch biosynthesis were positionally associated with significant RHM regions in white and yellow cassava populations. The hide-a-SNP analysis performed to validate the RHM results indicates that spurious associations due to linkage may have been avoided in the RHM analysis even when large segments were involved (Figure 2B). Given the genetic complexity of the cassava DM trait, we suggest that candidate genes, including invertases (neutral and acid) and FDA, may be targeted for gene editing or transgenic techniques to

substantiate the role of these genes in DM and starch accumulation in cassava and to provide a clear path for their utilization in cassava breeding programs.

DM content must work together with fresh root yield (FYLD) to make cassava production profitable and provide value for farmers and processors. To investigate whether some of the genes and gene families identified in the RHM analysis are also involved in the biological processes that lead to cassava FYLD, we validated their effects on FYLD using the same validation procedures and populations as above. The results showed prediction accuracies for SnRKs on FYLD as 0.03 (0.02), 53 likely candidates as 0.02 (0.02), 53 unlikely candidates as 0.006 (0.03), RHM-region genes as 0.03 (0.02), and cassava starch pathway genes as -0.009 (0.02). These results suggest no single biological pathway controls DM and FYLD. This is not surprising since there is little genetic correlation between DM and FYLD (Kawano et al., 1987). It appears from the negative correlation between carotenoid content in roots and DM content in African cassava germplasm (Esuma et al., 2016; Akinwale et al., 2010) and from the link between bHLH and sugar biosynthesis (Krügel et al., 2012), that yellow flesh color is associated with the accumulation of reducing sugars in edible roots (Eleazu and Eleazu, 2012). This poses a more complex challenge for improving DM in African yellow cassava and shifts attention towards finding recombinant yellow cassava progenies that have high DM. Ceballos et al. (2015) states that the search for the appropriate recombinant is difficult in cassava breeding and advocates for the use of inbred progenitors while breeding for hybrid cassava.

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

In this paper, we have utilized candidate gene analysis attempting to understand the function of the genes or gene families positionally associated with the RHM hits. We do not make the claim that these candidates are causal genes detected by the RHM hits but rather we have shown using prediction accuracies that these RHM hit loci were positionally associated with the DM trait in cassava (Figures 1, 4A and 4B) thus resulting in better predictability than random genes used as controls. To validate the hypotheses presented in this paper regarding candidate genes underlying DM accumulation in cassava, and to elucidate the physiological mechanisms involved in the expression of the DM trait in both yellow and white cassava, we recommend the use of genome editing and/or transgenic technology, and in-depth analysis of sugars and carbohydrates in cassava roots, stems and leaves. Similar studies in potato have benefited and informed potato breeding, and the same will be true of cassava as new insights become available.

Conclusion:

Using RHM analysis, we demonstrate the complex genetic architecture of DM content in high DM white African cassava. Candidate gene analysis revealed possible roles of SnRKs, vacuolar and neutral invertases, phosphoglucose isomerase and FDA in the regulation of sugar and starch biosynthesis in cassava. The RHM analysis indicated that inheritance of DM content in the high DM white cassava population is more polygenic than in the low DM yellow cassava population. We examined the utility of models based on genome-wide candidate genes found in this study using prediction accuracies in a different but related population and found appreciable predictive ability compared to what is obtained when whole genome markers were used. Transcriptional regulators such as bHLH may be involved in flesh root color and sugar biosynthesis in cassava, as shown in potato. We recommend further studies using genome editing or transgenic technology to better understand these mechanisms and to inform and accelerate breeding efforts for cassava.

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Declarations:

Author's contributions:

UGO designed, carried out study and drafted manuscript, DA provided statistical assistance and advice, JLJ supervised the study, designed the validation procedures and revised manuscript, IR and PK supervised data generation for this study and revised manuscript. All authors read and approved manuscript.

Acknowledgements:

We acknowledge the Bill & Melinda Gates Foundation and UKaid (Grant 1048542; http://www.gatesfoundation.org) and support from the CGIAR Research Program on Roots, Tubers and Bananas (http://www.rtb.cgiar.org). Deniz Akdemir was supported by the USDA-NIFA-AFRI Triticeae Coordinated Agricultural Project, award number 2011-68002-30029. We give special thanks to Professor Susan McCouch for her technical review and insights. Thanks also to A. I. Smith and technical teams at IITA for collection of other phenotypic data and to A. Agbona and M. Wolfe for data curation.

Competing interests

The authors declare that they have no competing interests.

References:

Adewusi, S.R. and Bradbury, J.H., 1993. Carotenoids in cassava: Comparison of open column and HPLC methods of analysis. Journal of the Science of Food and Agriculture, 62(4), pp.375-383.

Akinwale, M.G., Aladesanwa, R.D., Akinyele, B.O., Dixon, A.G.O. and Odiyi, A.C., 2010. Inheritance of-carotene in cassava (Manihot esculenta crantza). International Journal of Genetics and Molecular Biology, 2(10), pp.198-201.

Aniedu, C. and Omodamiro, R.M., 2012. Use of Newly Bred B-Carotene Cassava in Production of Value-Added Products: Implication for Food Security in Nigeria. Global Journal of Science Frontier Research, 12(10-D).

Ap Rees, T., 1992. Synthesis of storage starch. Carbon partitioning within and between organisms. Bios Scientific Publishers, Oxford, pp.115-131.

Appeldoorn, N.J., de Bruijn, S.M., Koot-Gronsveld, E.A., Visser, R.G., Vreugdenhil, D. and van der Plas, L.H., 1997. Developmental changes of enzymes involved in conversion of sucrose to hexose-phosphate during early tuberisation of potato. Planta, 202(2), pp.220-226.

Atmodjo, M.A., Sakuragi, Y., Zhu, X., Burrell, A.J., Mohanty, S.S., Atwood, J.A., Orlando, R., Scheller, H.V. and Mohnen, D., 2011. Galacturonosyltransferase (GAUT) 1

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

and GAUT7 are the core of a plant cell wall pectin biosynthetic homogalacturonan: galacturonosyltransferase complex. Proceedings of the National Academy of Sciences, 108(50), pp.20225-20230.

Ayres, D.L., Darling, A., Zwickl, D.J., Beerli, P., Holder, M.T., Lewis, P.O., Huelsenbeck, J.P., Ronquist, F., Swofford, D.L., Cummings, M.P. and Rambaut, A., 2011. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. Systematic biology, p.syr100.

Bahaji, A., Sánchez-López, Á.M., De Diego, N., Muñoz, F.J., Baroja-Fernández, E., Li, J., Ricarte-Bermejo, A., Baslam, M., Aranjuelo, I., Almagro, G. and Humplík, J.F., 2015. Plastidic phosphoglucose isomerase is an important determinant of starch accumulation in mesophyll cells, growth, photosynthetic capacity, and biosynthesis of plastidic cytokinins in Arabidopsis. PloS one, 10(3), p.e0119641.

Barrios, E.A., and R. Bressani. 1967. Composicion quimica de la raiz y de la hoja de algunas variedades de yuca Manihot. Turrialba 17:314-320.

Barry, G.F., Cheikh, N. and Kishore, G.M., Monsanto Technology Llc, 2002. Expression of fructose 1, 6 bisphosphate aldolase in transgenic plants. U.S. Patent 6,441,277.

Bland, J.M. and Altman, D.G., 1995. Multiple significance tests: the Bonferroni method. Bmj, 310(6973), p.170.

Bodmer, W. and Tomlinson, I., 2010. Rare genetic variants and the risk of cancer. Current opinion in genetics & development, 20(3), pp.262-267.

Bouis, H.E., Hotz, C., McClafferty, B., Meenakshi, J.V. and Pfeiffer, W.H., 2011. Biofortification: a new tool to reduce micronutrient malnutrition. Food and nutrition bulletin, 32(1 suppl1), pp.S31-S40.

Bredeson, J.V., Lyons, J.B., Prochnik, S.E., Wu, G.A., Ha, C.M., Edsinger-Gonzales, E., Grimwood, J., Schmutz, J., Rabbi, I.Y., Egesi, C. and Nauluvula, P., 2016. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity.Nature biotechnology, 34(5), pp.562-570.

Brummell, D.A., Chen, R.K., Harris, J.C., Zhang, H., Hamiaux, C., Kralicek, A.V. and McKenzie, M.J., 2011. Induction of vacuolar invertase inhibitor mRNA in potato tubers contributes to cold-induced sweetening resistance and includes spliced hybrid mRNA variants. Journal of experimental botany, 62(10), pp.3519-3534.

Burrell, M.M., Mooney, P.J., Blundy, M., Carter, D., Wilson, F., Green, J., Blundy, K.S. and Rees, T.A., 1994. Genetic manipulation of 6-phosphofructokinase in potato tubers. Planta, 194(1), pp.95-101.

Caballero, A., Tenesa, A. and Keightley, P.D., 2015. The Nature of Genetic Variation for Complex Traits Revealed by GWAS and Regional Heritability Mapping Analyses. Genetics, 201(4), pp.1601-1613.

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B., 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic acids research, 37(suppl 1), pp.D233-D238.

Ceballos, H., Iglesias, C.A., Pérez, J.C. and Dixon, A.G., 2004. Cassava breeding: opportunities and challenges. Plant molecular biology, 56(4), pp.503-516.

Ceballos, H., Kawuki, R.S., Gracen, V.E., Yencho, G.C. and Hershey, C.H., 2015. Conventional breeding, marker-assisted selection, genomic selection and inbreeding in clonally propagated crops: a case study for cassava. Theoretical and Applied Genetics, 128(9), pp.1647-1667.

Ceballos, H., Morante, N., Sánchez, T., Ortiz, D., Aragon, I., Chávez, A.L., Pizarro, M., Calle, F. and Dufour, D., 2013. Rapid cycling recurrent selection for increased carotenoids content in cassava roots. Crop Science, 53(6), pp.2342-2351.

Chávez, A.L., Sánchez, T., Jaramillo, G., Bedoya, J., Echeverry, J., Bolaños, E.A., Ceballos, H. and Iglesias, C.A., 2005. Variation of quality traits in cassava roots evaluated in landraces and improved clones.Euphytica, 143(1-2), pp.125-133.

Chincinska, I.A., Liesche, J., Krügel, U., Michalska, J., Geigenberger, P., Grimm, B. and Kühn, C., 2008. Sucrose transporter StSUT4 from potato affects flowering, tuberization, and shade avoidance response. Plant Physiology, 146(2), pp.515-528.

Crozet, P., Margalha, L., Confraria, A., Rodrigues, A., Martinho, C., Adamo, M., Elias, C.A. and Baena-González, E., 2014. Mechanisms of regulation of SNF1/AMPK/SnRK1 protein kinases. Frontiers in plant science, 5, p.190.

de Godoy, F., Bermúdez, L., Lira, B.S., de Souza, A.P., Elbl, P., Demarco, D., Alseekh, S., Insani, M., Buckeridge, M., Almeida, J. and Grigioni, G., 2013. Galacturonosyltransferase 4 silencing alters pectin composition and carbon partitioning in tomato. Journal of experimental botany, 64(8), pp.2449-2466.

De Jong, W.S., Eannetta, N.T., De Jong, D.M. and Bodis, M., 2004. Candidate gene analysis of anthocyanin pigmentation loci in the Solanaceae.Theoretical and Applied Genetics, 108(3), pp.423-432.

Deniz Akdemir and Okeke Uche Godfrey (2014). EMMREML: Fitting mixed models with known covariance structures. R package version 2.0. http://CRAN.R-project.org/package=EMMREML

Dietz, K.J., 1985. A possible rate-limiting function of chloroplast hexosemonophosphate isomerase in starch synthesis of leaves. Biochimica et Biophysica Acta (BBA)-General Subjects, 839(3), pp.240-248.

Du, J., Yin, H., Zhang, S., Wei, Z., Zhao, B., Zhang, J., Gou, X., Lin, H. and Li, J., 2012. Somatic embryogenesis receptor kinases control root development mainly via brassinosteroid□independent actions in Arabidopsis thaliana. Journal of integrative plant biology, 54(6), pp.388-399.

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Eleazu, C.O. and Eleazu, K.C., 2012. Determination of the proximate composition, total carotenoid, reducing sugars and residual cyanide levels of flours of 6 new yellow and white cassava (Manihot esculenta Crantz) varieties. American Journal of Food Technology, 7(10), pp.642-649.

El-Sharkawy, M.A., 2003. Cassava biology and physiology. Plant molecular biology, 53(5), pp.621-641.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one, 6(5), p.e19379.

EMBL-EBI InterPro. Protein Sequence analysis and classification. http://www.ebi.ac.uk/interpro/protein/D1G5D2/similar-proteins;jsessionid=10EFEE974F312EC7435985662EB07A7D. Accessed October 9, 2016.

Enidiok, S.E., Attah, L.E. and Otuechere, C.A., 2008. Evaluation of moisture, total cyanide and fiber contents of garri produced from cassava (Manihot utilissima) varieties obtained from Awassa in Southern Ethiopia.Pak. J. Nutr, 7, pp.625-629.

Esuma, W., Kawuki, R.S., Herselman, L. and Labuschagne, M.T., 2016. Diallel analysis of provitamin A carotenoid and dry matter content in cassava (Manihot esculenta Crantz). Breeding Science, 66(4), pp.627-635.

Food and Agriculture Organization of the United Nations. FAOSTAT Statistics Database. :FAO, 2013.

Geigenberger, P., 2003. Regulation of sucrose to starch conversion in growing potato tubers. Journal of Experimental Botany, 54(382), pp.457-465.

Gibson, G., 2012. Rare and common variants: twenty arguments. Nature Reviews Genetics, 13(2), pp.135-145.

Glaubitz, J., T. Casstevens, R. Elshire, J. Harriman, and E.S. Buckler. 2012. TASSEL 3.0 genotyping by sequencing (GBS) pipeline documentation. Edward S. Buckler, USDA-ARS, Ithaca, NY. http://www.maizegenetics.net/tassel/docs/TasselPipelineGBS.pdf (accessed 3. Jan. 2014)

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., ... & Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic acids research, 40(D1), D1178-D1186. (http://www.phytozome.net; accessed 1 July, 2014).

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Haake, V., Zrenner, R., Sonnewald, U. and Stitt, M., 1998. A moderate decrease of plastid aldolase activity inhibits photosynthesis, alters the levels of sugars and starch, and inhibits growth of potato plants. The Plant Journal, 14(2), pp.147-157.

Halford, N.G. and Hardie, D.G., 1998. SNF1-related protein kinases: global regulators of carbon metabolism in plants?. Plant molecular biology, 37(5), pp.735-748.

Halford, N.G., Hey, S., Jhurreea, D., Laurie, S., McKibbin, R.S., Paul, M. and Zhang, Y., 2003. Metabolic signalling and carbon partitioning: role of Snf1□related (SnRK1) protein kinase. Journal of Experimental Botany, 54(382), pp.467-475.

Heuer, B., Hansen, M.J. and Anderson, L.E., 1982. Light modulation of phosphofructokinase in pea leaf chloroplasts. Plant physiology, 69(6), pp.1404-1406.

Holden, H.M., Rayment, I. and Thoden, J.B., 2003. Structure and function of enzymes of the Leloir pathway for galactose metabolism. Journal of Biological Chemistry, 278(45), pp.43885-43888.

Holleman, L.W.J., and A. Aten. 1956. Elaboracion de la yuca y sus productos en las industrias rurales. Cuaderno de Fomento Agropecuario. Organizacion de las Naciones Unidas para la Agricultura y la Alimentacion. Bol. 54.

Iglesias, C.A. and Hershey, C.H., 1991, October. True cassava seed: research for a production alternative. In Symposium on Tropical Root Crops in a Developping Economy 380 (pp. 164-171).

John D. Storey with contributions from Andrew J. Bass, Alan Dabney and David Robinson (2015). qvalue: Q-value estimation for false discovery rate control. R package version 2.2.2. http://github.com/jdstorey/qvalue (Accessed May 5, 2016).

Jossier, M., Bouly, J.P., Meimoun, P., Arjmand, A., Lessard, P., Hawley, S., Grahame Hardie, D. and Thomas, M., 2009. SnRK1 (SNF1□related kinase 1) has a central role in sugar and ABA signalling in Arabidopsis thaliana. The Plant Journal, 59(2), pp.316-328.

Junker, B.H., 2004. Sucrose breakdown in the potato tuber. Mathematisch-Naturwissenschaftliche Fakultät. Universität Potsdam, 126.

Kawano, K., Fukuda, W.M.G. and Cenpukdee, U., 1987. Genetic and environmental effects on dry matter content of cassava root. Crop Science, 27(1), pp.69-74.

Kawano, K., Fukuda, W.M.G. and Cenpukdee, U., 1987. Genetic and environmental effects on dry matter content of cassava root. Crop Science,27(1), pp.69-74.

Kawano, K., Narintaraporn, K., Narintaraporn, P., Sarakarn, S., Limsila, A., Limsila, J., Suparhan, D., Sarawat, V. and Watananonta, W., 1998. Yield improvement in a multistage breeding program for cassava. Crop Science, 38(2), pp.325-332.

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Keating, B.A., Wilson, G.L. and Evenson, J.P., 1988. Effects of length, thickness, orientation, and planting density of cassava (Manihot esculenta Crantz) planting material on subsequent establishment, growth and yield. E. Afr. Agric. For. J, 53, pp.145-149.

Kimura, M., Kobori, C.N., Rodriguez-Amaya, D.B. and Nestel, P., 2007. Screening and HPLC methods for carotenoids in sweetpotato, cassava and maize for plant breeding trials. Food Chemistry, 100(4), pp.1734-1746.

Kleczkowski, L.A., 1999. A phosphoglycerate to inorganic phosphate ratio is the major factor in controlling starch levels in chloroplasts via ADP□glucose pyrophosphorylase regulation. FEBS letters, 448(1), pp.153-156.

Kleczkowski, L.A., 2000. Is leaf ADP-glucose pyrophosphorylase an allosteric enzyme?. Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology, 1476(1), pp.103-108.

Klein, R.J., 2007. Power analysis for genome-wide association studies. BMC genetics, 8(1), p.58.

Krügel, U., He, H.X., Gier, K., Reins, J., Chincinska, I., Grimm, B., Schulze, W.X. and Kühn, C., 2012. The potato sucrose transporter StSUT1 interacts with a DRM-associated protein disulfide isomerase. Molecular plant, 5(1), pp.43-62.

La Frano, M.R., Woodhouse, L.R., Burnett, D.J. and Burri, B.J., 2013. Biofortified cassava increases β-carotene and vitamin A concentrations in the TAG-rich plasma layer of American women. British Journal of Nutrition,110(02), pp.310-320.

Laurie, S., McKibbin, R.S. and Halford, N.G., 2003. Antisense SNF1□related (SnRK1) protein kinase gene represses transient activity of an α□amylase (α□Amy2) gene promoter in cultured wheat embryos. Journal of Experimental Botany, 54(383), pp.739-747.

Li, B., Liu, H., Zhang, Y., Kang, T., Zhang, L., Tong, J., Xiao, L. and Zhang, H., 2013. Constitutive expression of cell wall invertase genes increases grain yield and starch content in maize. Plant biotechnology journal, 11(9), pp.1080-1091.

Lim, H.K. 1968. Composition data of feeds and concentrates. Malay.Agric.J.46:63-79.

Lin, Y., Liu, T., Liu, J., Liu, X., Ou, Y., Zhang, H., Li, M., Sonnewald, U., Song, B. and Xie, C., 2015. Subtle regulation of potato acid invertase activity by a protein complex of invertase, invertase inhibitor, and SUCROSE NONFERMENTING1-RELATED PROTEIN KINASE. Plant physiology,168(4), pp.1807-1819.

Liu, W., Zhou, Y., Sanchez, T., Ceballos, H. and White, W.S., 2010. The vitamin A equivalence of {beta}-carotene in {beta}-carotene-biofortified cassava ingested by women. The FASEB Journal, 24(1_MeetingAbstracts), pp.92-7.

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Long, A.D. and Langley, C.H., 1999. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. Genome Research, 9(8), pp.720-731.

Ly, Delphine, et al.; Relatedness and genotype× environment interaction affect prediction accuracies in genomic selection: A study in cassava." Crop Science53.4 (2013): 1312-1325.

Maass, D., Arango, J., Wüst, F., Beyer, P. and Welsch, R., 2009. Carotenoid crystal formation in Arabidopsis and carrot roots caused by increased phytoene synthase protein levels. PLoS One, 4(7), p.e6373.

Maziya-Dixon, B., A.G.O. Dixon, and A.-R.A. Adebowale. 2007. Targeting different end uses of cassava: Genotypic variations for cyanogenic potentials and pasting properties. Int. J. Food Sci. Technol. 42:969–976. doi:10.1111/j.1365-2621.2006.01319.x

Mckenzie, M.J., Chen, R.K., Harris, J.C., Ashworth, M.J. and Brummell, D.A., 2013. Post□translational regulation of acid invertase activity by vacuolar invertase inhibitor affects resistance to cold□induced sweetening of potato tubers. Plant, cell & environment, 36(1), pp.176-185.

McKibbin, R.S., Muttucumaru, N., Paul, M.J., Powers, S.J., Burrell, M.M., Coates, S., Purcell, P.C., Tiessen, A., Geigenberger, P. and Halford, N.G., 2006. Production of high□starch, low□glucose potatoes through over□expression of the metabolic regulator SnRK1. Plant biotechnology journal,4(4), pp.409-418.

Meenakshi, J.V., Banerji, A., Manyong, V., Tomlins, K., Hamukwala, P., Zulu, R. and Mungoma, C., 2010. Consumer acceptance of provitamin A orange maize in rural Zambia (HarvestPlus Working Paper No. 4).

Meenakshi, J.V., Johnson, N., Victor, M., De Groote, H., Javelosa, J., Yanggen, D., Naher, F., Garcia, J. and Meng, E., 2007. HarvestPlus Working Paper No. 2.

Miller, M.E. and Chourey, P.S., 1992. The maize invertase-deficient miniature-1 seed mutation is associated with aberrant pedicel and endosperm development. The Plant Cell, 4(3), pp.297-305.

Mitchell, A., Chang, H.Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S. and Sangrador-Vegas, A., 2014. The InterPro protein families database: the classification resource after 15 years. Nucleic acids research, p.gku1243.

Momma, M. and Fujimoto, Z., 2012. Interdomain disulfide bridge in the rice granule bound starch synthase I catalytic domain as elucidated by X-ray structure analysis. Bioscience, biotechnology, and biochemistry, 76(8), pp.1591-1595.

Nagamine, Y., Pong-Wong, R., Navarro, P., Vitart, V., Hayward, C., Rudan, I., Campbell, H., Wilson, J., Wild, S., Hicks, A.A. and Pramstaller, P.P., 2012. Localising

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

loci underlying complex trait variation using regional genomic relationship mapping. PloS one, 7(10), p.e46501.

Okechukwu, R.U., and A.G.O. Dixon. 2008. Genetic gains from 30 years of cassava breeding in Nigeria for storage root yield and disease resistance in elite cassava genotypes. J. Crop Improv. 22:181–208. doi:10.1080/15427520802212506

Pfeiffer, W.H. and McClafferty, B., 2007. HarvestPlus: breeding crops for better nutrition. Crop Science, 47(Supplement_3), pp.S-88.

Plus, H., 2009. Breeding crops for better nutrition. Provitamin A Cassava. http://r4d.dfid.gov.uk/PDF/Outputs/Misc_Crop/HarvstPlus_Cassava_Strategy.pdf. Accessed June 30, 2016.

Polge, C. and Thomas, M., 2007. SNF1/AMPK/SnRK1 kinases, global regulators at the heart of energy control?. Trends in plant science, 12(1), pp.20-28.

Purcell, P.C., Smith, A.M. and Halford, N.G., 1998. Antisense expression of a sucrose non□fermenting□1□related protein kinase sequence in potato results in decreased expression of sucrose synthase in tubers and loss of sucrose□inducibility of sucrose synthase transcripts in leaves. The Plant Journal,14(2), pp.195-202.

Quinlan, A.R. and Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26(6), pp.841-842.

R Development Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org (January, 2016).

Radchuk, R., Emery, R.J., Weier, D., Vigeolas, H., Geigenberger, P., Lunn, J.E., Feil, R., Weschke, W. and Weber, H., 2010. Sucrose non□fermenting kinase 1 (SnRK1) coordinates metabolic and hormonal signals during pea cotyledon growth and differentiation. The Plant Journal, 61(2), pp.324-338.

Radchuk, R., Radchuk, V., Weschke, W., Borisjuk, L. and Weber, H., 2006. Repressing the expression of the SUCROSE NONFERMENTING-1-RELATED PROTEIN KINASE gene in pea embryo causes pleiotropic defects of maturation similar to an abscisic acid-insensitive phenotype. Plant Physiology, 140(1), pp.263-278.

Raji, A.A., Dixon, A.G.O. and Ladeinde, T.A.O., 2007. Agronomic traits and tuber quality attributes of farmer grown cassava (Manihot esculenta) landraces in Nigeria. Journal of Tropical Agriculture, 45, pp.9-13.

Renz, A., Merlo, L. and Stitt, M., 1993. Partial purification from potato tubers of three fructokinases and three hexokinases which show differing organ and developmental specificity. Planta, 190(2), pp.156-165.

Resende, R.T., Resende, M.D.V., Silva, F.F., Azevedo, C.F., Takahashi, E.K., Silva-Junior, O.B. and Grattapaglia, D., 2017. Regional heritability mapping and genome-wide

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

association identify loci for complex growth, wood and disease resistance traits in Eucalyptus. *New Phytologist*, *213*(3), pp.1287-1300.

Riggio, V. and Pong-Wong, R., 2014, October. Regional Heritability Mapping to identify loci underlying genetic variation of complex traits. In BMC proceedings (Vol. 8, No. 5, p. 1). BioMed Central.

Riggio, V., Matika, O., Pong-Wong, R., Stear, M.J. and Bishop, S.C., 2013. Genome-wide association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs. Heredity, 110(5), pp.420-429.

Rolland, F., Moore, B. and Sheen, J., 2002. Sugar sensing and signaling in plants. The plant cell, 14(suppl 1), pp.S185-S205.

Safo□Kantanka, O. and Owusu□Nipah, J., 1992. Cassava varietal screening for cooking quality: relationship between dry matter, starch content, mealiness and certain microscopic observations of the raw and cooked tuber. Journal of the Science of Food and Agriculture, 60(1), pp.99-104.

Saithong, T., Rongsirikul, O., Kalapanulak, S., Chiewchankaset, P., Siriwat, W., Netrphan, S., Suksangpanomrung, M., Meechai, A. and Cheevadhanarak, S., 2013. Starch biosynthesis in cassava: a genome-based pathway reconstruction and its exploitation in data integration. BMC systems biology, 7(1), p.75.

Shirali, M., Pong-Wong, R., Navarro, P., Knott, S., Hayward, C., Vitart, V., Rudan, I., Campbell, H., Hastie, N.D., Wright, A.F. and Haley, C.S., 2015. Regional heritability mapping method helps explain missing heritability of blood lipid traits in isolated populations. Heredity.

Sorey, J.D. and Tibshirani, R., 2003. Statistical significance for genome-wide experiments. Proc. Natl Acad. Sci. USA, 100, pp.9440-9445.

Ssemakula, G., Dixon, A.G.O. and Maziya-Dixon, B., 2007. Stability of total carotenoid concentration and fresh yield of selected yellow-fleshed cassava (Manihot esculenta Crantz). Journal of Tropical Agriculture, 45(1-2), pp.14-20.

Sweetlove, L.J., Kruger, N.J. and Hill, S.A., 2001. Starch synthesis in transgenic potato tubers with increased 3-phosphoglyceric acid content as a consequence of increased 6-phosphofructokinase activity. Planta, 213(3), pp.478-482.

Tai, H.H., Goyer, C. and Murphy, A.M., 2013. Potato MYB and bHLH transcription factors associated with anthocyanin intensity and common scab resistance. Botany, 91(10), pp.722-730.

Takatsuji, H., 1998. Zinc-finger transcription factors in plants. Cellular and Molecular Life Sciences CMLS, 54(6), pp.582-596.

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Takatsuji, H., 1999. Zinc-finger proteins: the classical zinc finger emerges in contemporary plant science. Plant molecular biology, 39(6), pp.1073-1078.

Tanaka, M., Takahata, Y., Nakayama, H., Nakatani, M. and Tahara, M., 2009. Altered carbohydrate metabolism in the storage roots of sweetpotato plants overexpressing the SRF1 gene, which encodes a Dof zinc finger transcription factor. Planta, 230(4), pp.737-746.

Tang, X., Su, T., Han, M., Wei, L., Wang, W., Yu, Z., Xue, Y., Wei, H., Du, Y., Greiner, S. and Rausch, T., 2016. Suppression of extracellular invertase inhibitor gene expression improves seed weight in soybean (Glycine max). Journal of Experimental Botany, p.erw425.

Tiessen, A., Prescha, K., Branscheid, A., Palacios, N., McKibbin, R., Halford, N.G. and Geigenberger, P., 2003. Evidence that SNF1□related kinase and hexokinase are involved in separate sugar□signalling pathways modulating post□translational redox activation of ADP□glucose pyrophosphorylase in potato tubers. The Plant Journal, 35(4), pp.490-500.

Toledo-Ortiz, G., Huq, E. and Rodríguez-Concepción, M., 2010. Direct regulation of phytoene synthase gene expression and carotenoid biosynthesis by phytochrome-interacting factors. Proceedings of the National Academy of Sciences, 107(25), pp.11626-11631.

Uemoto, Y., Pong-Wong, R., Navarro, P., Vitart, V., Hayward, C., Wilson, J.F., Rudan, I., Campbell, H., Hastie, N.D., Wright, A.F. and Haley, C.S., 2013. The power of regional heritability analysis for rare and common variant detection: simulations and application to eye biometrical traits. Frontiers in genetics, 4, p.232.

VanRaden PM: Efficient methods to compute genomic predictions. J Dairy Sci. 2008, 91: 4414-4423. 10.3168/jds.2007-0980.

Vimala, B., Nambisan, B., Theshara, R. and Unnikrishnam, M., 2008. Variability of carotenoids in yellow-flesh cassava (Manihot esculatar Crantz).Geneconserve. Pro. br, 1.

Wang, E., Wang, J., Zhu, X., Hao, W., Wang, L., Li, Q., Zhang, L., He, W., Lu, B., Lin, H. and Ma, H., 2008. Control of rice grain-filling and yield by a gene with a potential signature of domestication. Nature genetics, 40(11), pp.1370-1374.

Wang, L. and Ruan, Y.L., 2012. New insights into roles of cell wall invertase in early seed development revealed by comprehensive spatial and temporal expression patterns of GhCWIN1 in cotton. Plant physiology, 160(2), pp.777-787.

Weschke, W., Panitz, R., Gubatz, S., Wang, Q., Radchuk, R., Weber, H. and Wobus, U., 2003. The role of invertases and hexose transporters in controlling sugar ratios in maternal and filial tissues of barley caryopses during early development. The Plant Journal, 33(2), pp.395-411.

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Weschke, W., Panitz, R., Sauer, N., Wang, Q., Neubohn, B., Weber, H. and Wobus, U., 2000. Sucrose transport into barley seeds: molecular characterization of two transporters and implications for seed development and starch accumulation. The Plant Journal, 21(5), pp.455-467.

Wolfe, M.D., Rabbi, I.Y., Egesi, C., Hamblin, M., Kawuki, R., Kulakow, P., Lozano, R., Carpio, D.P.D., Ramu, P. and Jannink, J.L., 2016. Genome-wide association and prediction reveals genetic architecture of cassava mosaic disease resistance and prospects for rapid genetic improvement. The plant genome, 9(2).

Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J.A., Kutalik, Z. and Amin, N., 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. Nature genetics, 46(11), pp.1173-1186.

Xue-Fei, D., Na, C., Li, W., Xiao-Cui, Z., Bo, Q., Tian-Lai, L. and Guo-Liang, Z., 2012. The SnRK Protein Kinase Family and the Function of SnRK1 Protein Kinase. International Journal of Agriculture & Biology, 14(4).

Yang, J., An, D. and Zhang, P., 2011. Expression profiling of cassava storage roots reveals an active process of glycolysis/gluconeogenesis. Journal of integrative plant biology, 53(3), pp.193-211.

Zanor, M.I., Osorio, S., Nunes-Nesi, A., Carrari, F., Lohse, M., Usadel, B., Kühn, C., Bleiss, W., Giavalisco, P., Willmitzer, L. and Sulpice, R., 2009. RNA interference of LIN5 in tomato confirms its role in controlling Brix content, uncovers the influence of sugars on the levels of fruit hormones, and demonstrates the importance of sucrose cleavage for normal fruit development and fertility. Plant physiology, 150(3), pp.1204-1218.

Zeng, Y., Navarro, P., Fernandez-Pujals, A.M., Hall, L.S., Clarke, T.K., Thomson, P.A., Smith, B.H., Hocking, L.J., Padmanabhan, S., Hayward, C. and MacIntyre, D.J., 2016. A Combined Pathway and Regional Heritability Analysis Indicates NETRIN1 Pathway is Associated with Major Depressive Disorder. Biological Psychiatry.

Zhang, Y., Jung, C.S. and De Jong, W.S., 2009. Genetic analysis of pigmented tuber flesh in potato. Theoretical and Applied Genetics, 119(1), pp.143-150.

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050

Figure 1: **Manhattan plots showing dry matter content genomic segment associations.** Upper and lower figures show RHM discovery associations for white and yellow cassava populations, respectively.

Figure 2A: **Genome-wide linkage disequilibrium between segments in the RHM analysis**. Linkage disequilibrium is measured as the mean correlation between all pairs of SNPs where one SNP is on one segment and the other is on the adjacent segment.

Figure 2B: **Histogram of the size of genomic segments in the RHM analysis**. The size of the window is the physical distance in base pairs between the first and the last of the 100 SNPs in the window.

Figure 3: **Sucrose/starch metabolism in a heterotrophic plant cell like the cassava tuber.** Key enzymes including sucrose transporter, invertase, phospho-glucose isomerase, and phosphofructokinase were within 500Kb of significant RHM segments.

Figure 4A: **Selected candidate genes and positions of significant RHM segments.** Circos plot of carbohydrate biosynthesis candidate genes or gene families and significant RHM segments shown by paired dotted lines. Points are randomly scattered along the y-axis to avoid overlaps and better visualize gene families.

Figure 4B: **Zoom-in plot of candidate genes and significant RHM segments in a 21Mb region of Chromosome 1.** The same genes or gene families as Fig. 4A are shown along with two significant RHM segments. The double line separates candidate genes with random y-axis positions from $-\log_{10}(\text{LFDR})$ plotting of the significance of RHM segments.

Table 1. Candidate genes and gene families associated with significant RHM regions.

RHM candidates for White DM											
Chr.	**Seg. tag SNP_ID**	**Seg. tag SNP bp**	**Seg. Span**	**LFDR**	**Target Seg.**	**Genomic**	**Candidate Genes**	**Phyt. ID**	**start (bp)**	**end (bp)**	**homologs**
1	S1_3416180	3416180	3064700 - 3416180	0.047227533	0.66	0.84	BAK1-interacting receptor-like kinase 1 (BIR1)	Manes.01G018900	3095915	3098209	AT5G48380.1
							BAK1-interacting receptor-like kinase 1 (BIR1)	Manes.01G019000	3132412	3134588	AT5G48380.1
1	S1_19936380	19936380	19646964 - 19936380	0.030168971	0.58	0.82	UDP-D-glucuronate 4-epimerase 2	Manes.01G074000	19949560	19951824	AT1G02000.1
							DHHC-type zinc finger family protein	Manes.01G073100	19818085	19825785	AT3G60800.1
							Bifunctional inhibitor trypsin-alpha amylase inhibitor	Manes.01G074600	20036878	20037858	AT1G62790.2
							Plant invertase/pectin methylesterase inhibitor superfamily protein	Manes.01G074800	20042476	20043982	AT4G25260.1
							Pectinesterase	Manes.01G075600	20158111	20159120	AT5G07420.1
							beta vacuolar processing enzyme	Manes.01G075700	20173648	20176768	AT1G62710.1
							mini zinc finger 2	Manes.01G090800	21547715	21548005	AT3G28917.1
							galactosyltransferase1	Manes.01G091600	21599298	21603804	AT1G26810.1
							starch synthase 2 (ADP-Glucose type)	Manes.01G091700	21623316	21629007	AT3G01180.1
							cycling DOF factor 2	Manes.01G092100	21647142	21650438	AT5G39660.1
4	S4_537966	537966	510971 - 732069	0.003820074	0.67	0.84	serine/threonine-protein kinase RIO	Manes.04G004800	535401	538252	AT1G08290.1
							Glycosyltransferase family 29	Manes.04G004900	542055	543236	AT1G08280.1
4	S4_859155	859155	623461 - 859155	0.00113334	0.62	0.82	alkaline/neutral invertase	Manes.04G006900	778931	784740	AT5G22510.1
							cellulose synthase-like A02	Manes.04G009400	1064431	1069211	AT5G22740.1
							C2H2-like zinc finger protein	Manes.04G008800	962985	963932	AT4G35280.1
							serine/threonine protein kinase (SnRK1)	Manes.04G006600	757312	760778	AT3G44610.1
5	S5_7279374	7279374	7061162 - 7279374	0.03310678	0.69	0.85	trehalose-phosphatase/synthase 9	Manes.05G087900	6905372	6911159	AT1G23870.1
							Serine/threonine-protein kinase WNK (With No Lysine)-related	Manes.05G089000	7284514	7286444	AT1G60060.1
							serine/threonine protein kinase (SnRK1)	Manes.05G090100	7365593	7370118	AT1G24030.2
							NAD(P)-linked oxidoreductase superfamily protein	Manes.05G092400	7630814	7632642	AT1G59960.1
							galactosyltransferase family protein	Manes.05G095000	7849628	7853700	AT5G62620.1
							bHLH Transcription factor	Manes.05G094700	7824312	7824497	LOC_Os07g09590.1
							Galactose oxidase	Manes.05G096300	7992476	7993801	Cre06.g306000.t1.1
							alpha-amylase-like 3	Manes.05G097100	8094930	8103684	AT1G69830.1
							sucrose transport protein (SUC3)	Manes.05G099000	8344933	8345349	LOC_Os02g36700.1
							UDP-galactose/UDP-glucose transporter	Manes.05G101600	8633269	8636565	Potri.016G139100.1

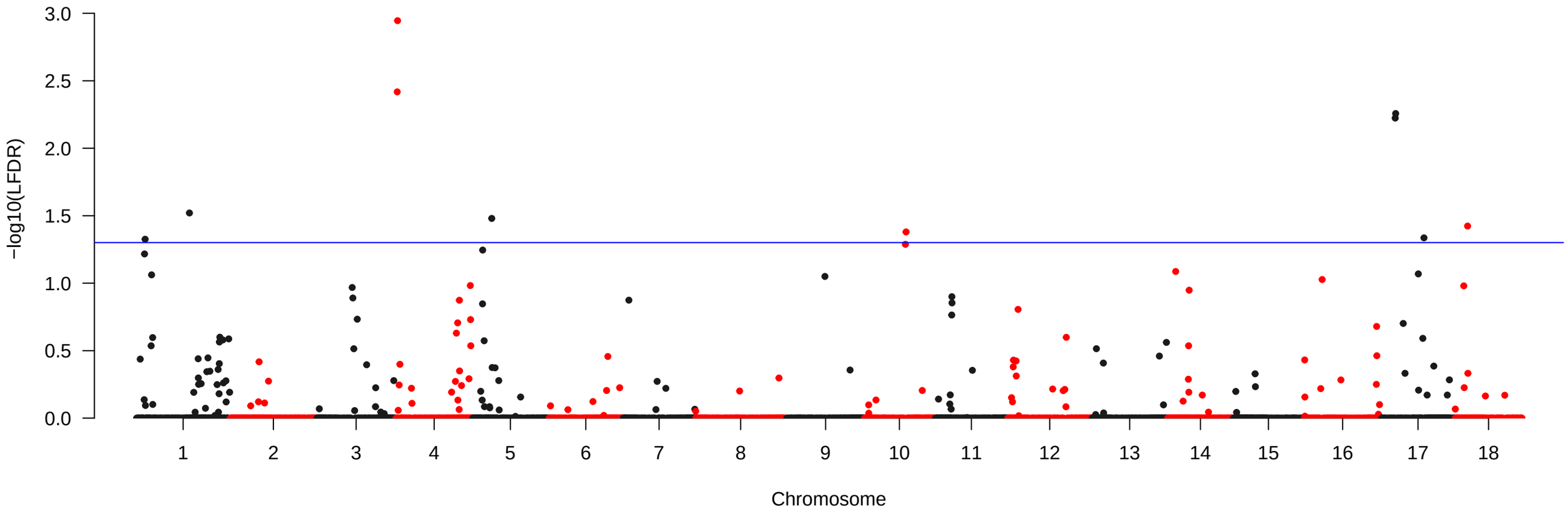
RHM candidates for White DM											
Chr.	**Seg. tag SNP_ID**	**Seg. tag SNP bp**	**Seg. Span**	**LFDR**	**Target Seg.**	**Genomic**	**Candidate Genes**	**Phyt. ID**	**start (bp)**	**end (bp)**	**homologs**
							zinc finger, C3HC4 type domain containing protein	Manes.05G102000	8696530	8696856	LOC_Os03g20870.1
10	S10_15606779	15606779	15024128 - 15606779	0.041702151	0.15	0.64	serine/threonine protein kinase (SnRK1)	Manes.10G079500	14844954	14851620	AT1G49730.1
							basic helix-loop-helix (bHLH) DNA-binding superfamily protein	Manes.10G080300	15279304	15281955	AT1G49770.1
							salt tolerance zinc finger	Manes.10G080000	15071612	15072680	AT1G27730.1
							Protein with RNI-like/FBD-like domains	Manes.10G080200	15240978	15241397	AT5G56410.1
							Regulator of Vps4 activity in the MVB pathway protein	Manes.10G080400	15346375	15347741	AT2G14830.2
17	S17_5368263	5368263	5152794 - 5368263	0.005975712	0.17	0.64	xyloglucan endotransglucosylase/hydrolase 30	Manes.17G015100	5198786	5201174	AT1G32170.1
17	S17_5532660		5285956 - 5532660	0.005529599	0.27	0.67	Pectin lyase-like superfamily protein	Manes.17G015200	5264305	5276255	AT3G07840.1
							Basic helix-loop-helix (bHLH) DNA-binding family protein	Manes.17G016000	5657214	5660207	AT1G32640.1
17	S17_16084946	16084946	15509115 - 16084946	0.046144077	0.14	0.62	CBL-interacting protein kinase 23 (SnRK3)	Manes.17G073900	21282260	21287846	AT1G30270.1
							inositol transporter 2	Manes.17G073000	21211865	21215459	AT1G30220.1
							Transducin/WD40 repeat-like superfamily protein	Manes.17G073300	21251484	21254389	AT1G65030.1
18	S18_4988720	4988720	4586953 - 4988720	0.037726745	0.69	0.85	sucrose transporter 4	Manes.18G054200	4548075	4559586	AT1G09960.1
							Nucleotide-diphospho-sugar transferases superfamily protein	Manes.18G054400	4572586	4576353	AT1G27600.1
							beta HLH protein 93	Manes.18G055000	4655203	4657015	AT5G65640.1
							CBL-interacting protein kinase 8 (SnRK3)	Manes.18G055300	4677895	4683026	AT4G24400.1
							UDP-Glycosyltransferase superfamily protein	Manes.18G056200	4793379	4795015	AT5G49690.1
							GATA-type zinc finger protein with TIFY domain	Manes.18G056300	4820431	4826234	AT4G24470.3
RHM candidates Yellow DM											
1	S1_22566390	22566390	21910139 - 22566390	0.04682523	0.72	0.87	galacturonosyltransferase 13	Manes.01G098400	22245408	22256933	AT3G01040.1
							Plant calmodulin-binding protein-related	Manes.01G099000	22282570	22286073	AT5G39380.1
							UDP-glucosyl transferase 76E11	Manes.01G100300	22370079	22371643	AT3G46670.1
							SNF1 kinase (SnRK1)	Manes.01G100900	22409977	22417397	AT3G01090.2
							Vesicle transport v-SNARE family protein	Manes.01G101200	22421446	22425558	AT5G39510.1
							brassinosteroid-6-oxidase 2	Manes.01G102800	22514594	22518856	AT3G30180.1
*9	S9_13325846	13325846	12810619 - 13747638		0.26	0.66	phosphofructokinase 2 (PFK2)	Manes.09G077800	13223145	13226580	AT5G47810.1

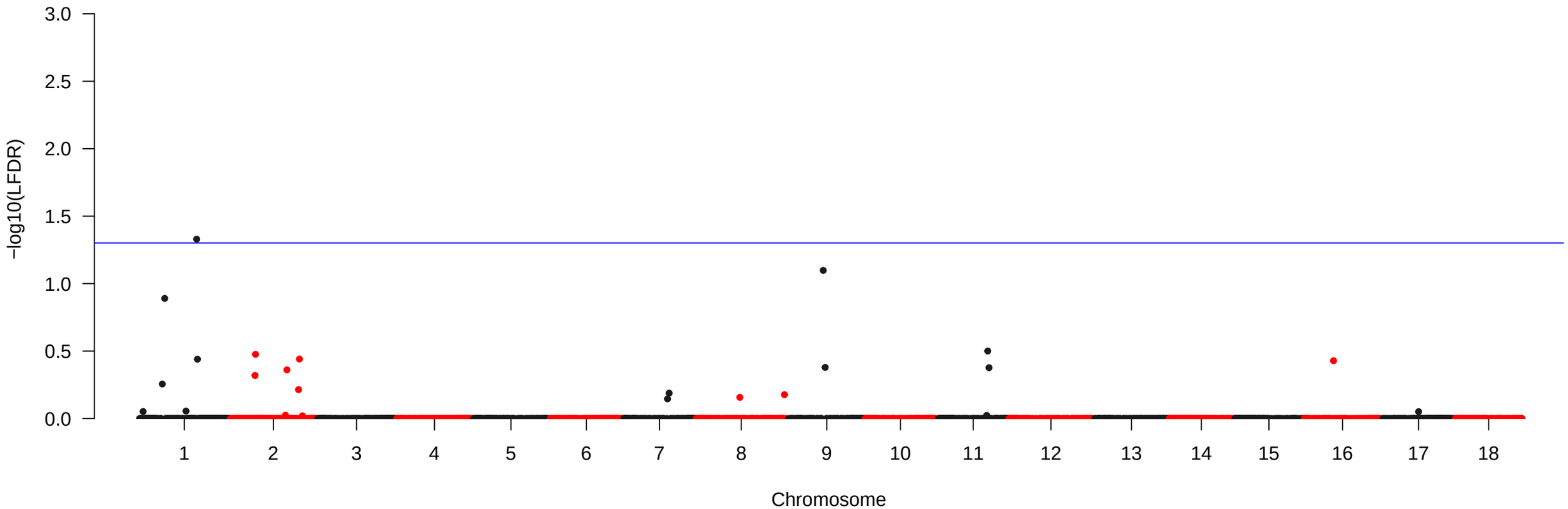
Chr.: chromosome, Seg.: segment, Phyt.: Phytozome.
*: Strong but non-significant signal

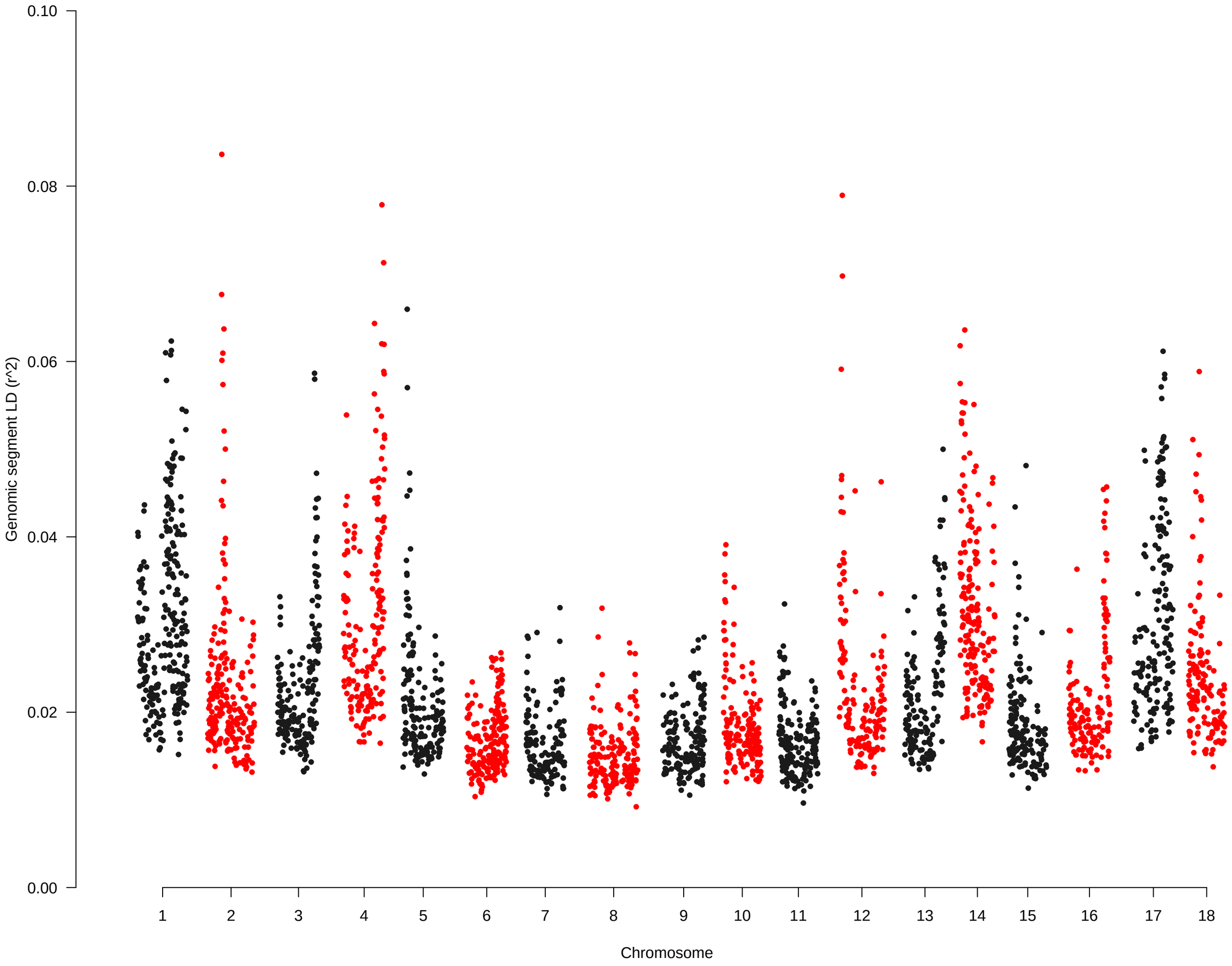
Table 2: Summary of validation results for RHM significant candidates. Prediction accuracies from selected candidate genes or genomic segments used to validate significance of the RHM hits are given with the number of markers used in each analysis (left column) or the standard deviation of cross validation repeat cycles (right column) in parentheses.

Genomic Segment	**Prediction accuracy**
SnRKs (7,203)	0.26 (0.04)
Random control for SnRKs (7,203)	0.12 (0.06)
53 Likely candidates (400)	0.17 (0.03)
53 Unlikely candidates (395)	0.14 (0.02)
Random control for 53 candidates (402)	0.06 (0.08)
RHM-region genes (2,297)	0.17 (0.04)
Cassava Starch genes (419)	0.18 (0.03)
Random-650 (2,300)	0.03 (0.01)
Whole genome SNPs (177,201)	0.54 (0.03)

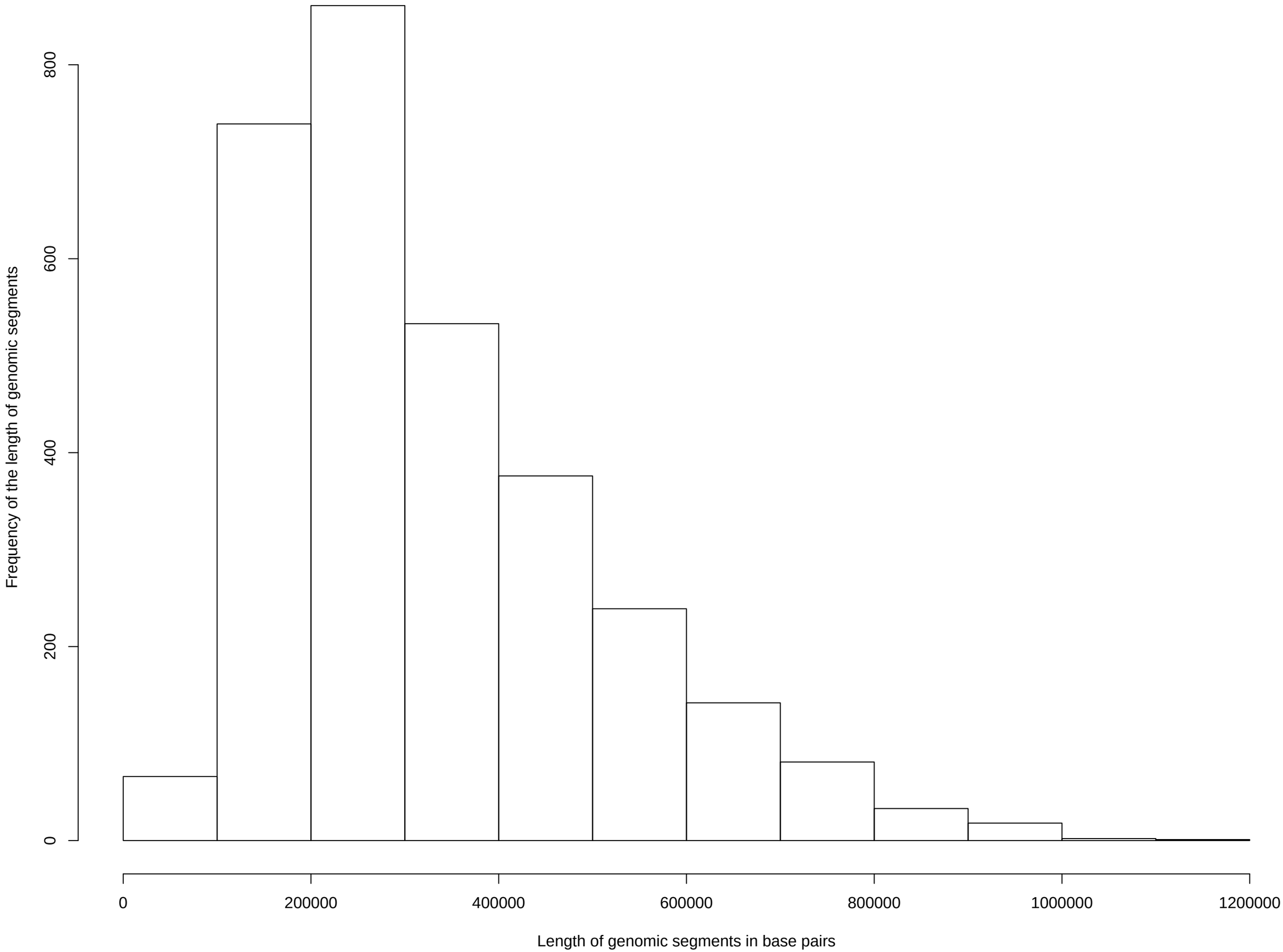
Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050






0.10
0.08
0.06
0.04
0.02
0.00
Genomic segment LD (r^2)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Chromosome

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050



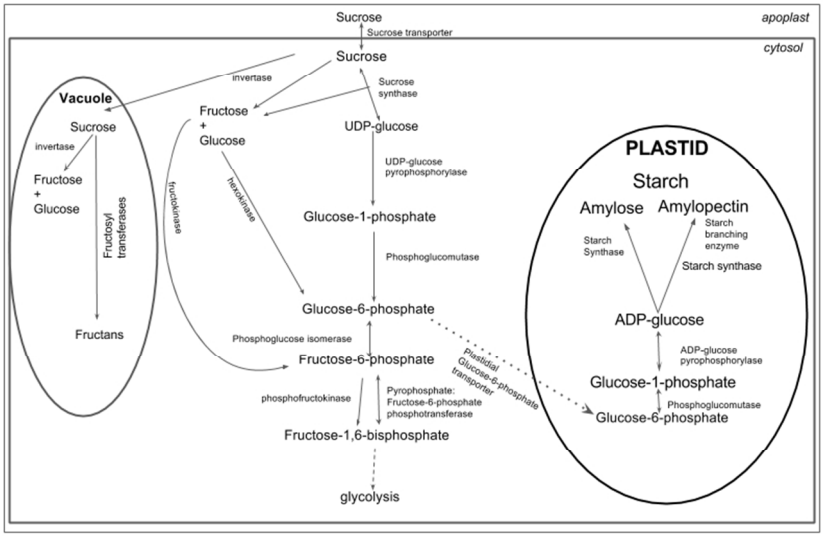
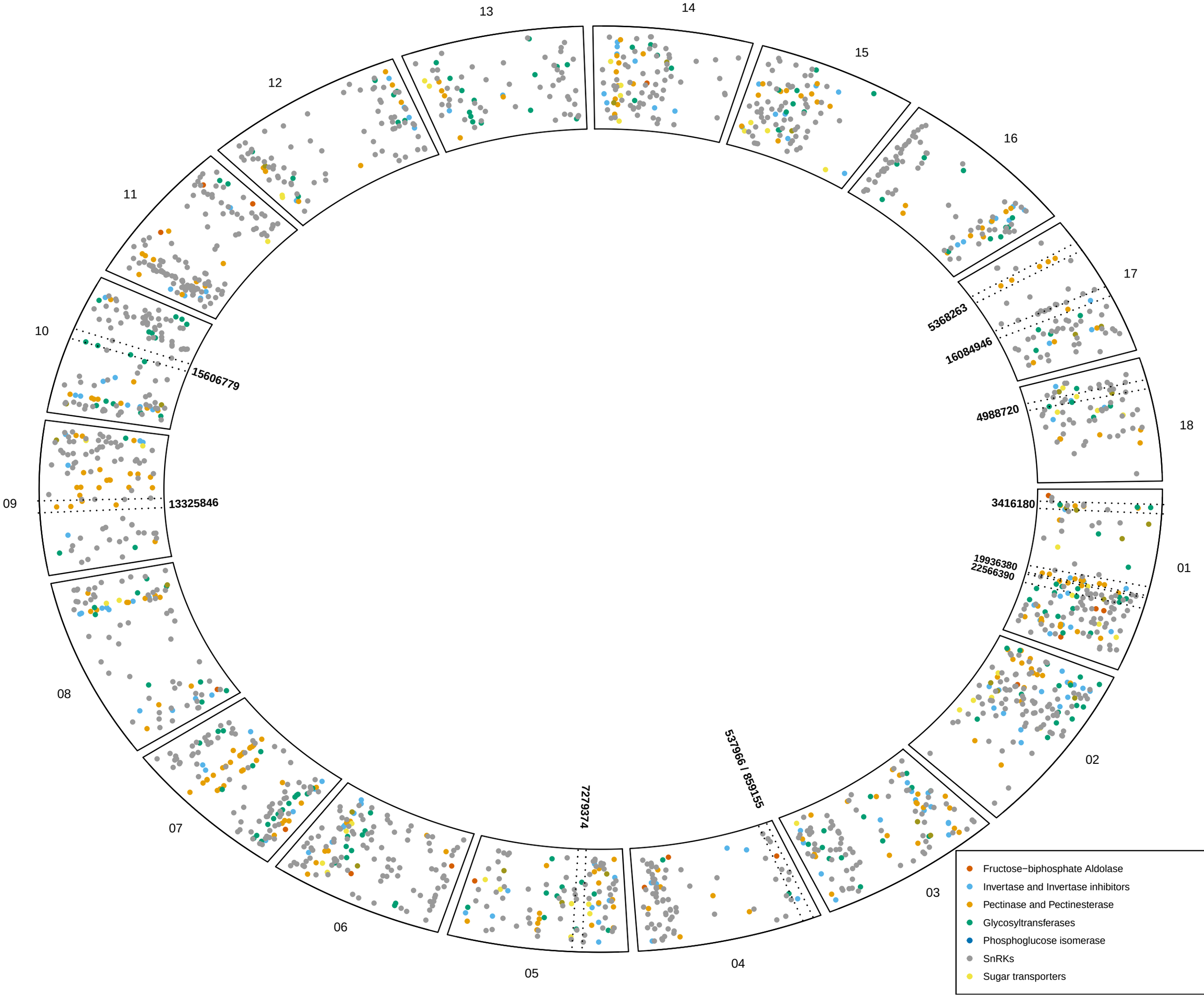


Figure 3

338x190mm (72 x 72 DPI)

Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050



Plant Gen. Accepted Paper, posted 07/31/2017. doi:10.3835/plantgenome2017.06.0050