CrossMark

Journal of Agricultural Science (2017), **155**, 1424–1441. © Cambridge University Press 2017
doi:10.1017/S0021859617000612

CROPS AND SOILS RESEARCH PAPER

Genome-wide association study for resistance to cassava root rot

A. C. BRITO[1], S. A. S. OLIVEIRA[1,2] AND E. J. OLIVEIRA[1,2]*

[1] *Embrapa Mandioca e Fruticultura, Rua da Embrapa, Caixa Postal 007, Zip Code 44380-000 Cruz das Almas, BA, Brazil*
[2] *Universidade Federal do Recôncavo da Bahia, Campus Cruz das Almas, Zip Code 44380-000 Cruz das Almas, BA, Brazil*

(*Received 23 January 2017; revised 24 August 2017; accepted 27 September 2017; first published online 26 October 2017*)

SUMMARY

Cassava root rot (CRR) disease associated with a complex of soil pathogens has caused great yield losses in the crop. The objective of the current work was to obtain insights about the genetic architecture of CRR resistance caused by *Fusarium* (dry root rot – DRR), *Phytophthora* (soft root rot – SRR) and Botryosphaeriaceae (black root rot – BRR) species, using genome-wide association studies (GWAS). Phenotyping data of 263 accessions (artificial inoculation) and 14 094 single-nucleotide polymorphisms (SNP) (missing data <0·10 and minor allele frequency >0·05) were used. The severity of CRR in peel and pulp was variable among accessions, but the pathogens that caused DRR were more aggressive. Broad-sense heritability (h_g^2) was of medium magnitude for all groups of resistances for pathogens, with variation from 0·16 ± 0·019 (Fspp Pulp) to 0·31 ± 0·028 (Fspp Peel). The kinship matrix was used to correct for stratification as well as for clustering the accessions. Overall, this analysis showed that there was no relationship between agronomic traits and resistance to CRR and the four clusters obtained from kinship matrix. The GWAS identified 38 significant SNPs, of which eight and 22 are related to the severity of DRR in the pulp and peel, respectively. The other eight SNPs were associated with SRR-peel (1), SRR-pulp (1), BRR-peel (3) and BRR-pulp (3). Half of the SNPs associated with CRR resistance have functional annotations related to defence and response to pathogen attack as well as general cellular processes. The current study revealed that resistance to CRR is controlled by multiple loci with small effects, and significant SNPs can be used to identify putative genes that control these traits.

INTRODUCTION

Cassava (*Manihot esculenta* Crantz) is one of the most important food sources in the world, providing energy to around 800 million people, mainly in the tropics (Ceballos *et al*. 2010; FAO 2013; Khumaida *et al*. 2015). The species can be adapted to adverse climatic conditions, especially in regions with low rainfall levels and low fertility soils, thus making the crop essential for communities living in these regions (Nwokoro *et al*. 2002) and serving as an excellent agricultural alternative to meet the growing global demand for food (Soto *et al*. 2015).

Cassava is produced throughout Brazil but mainly in the north and northeast regions, using different crop production systems and landraces, which generally present low yield potential. One of the most surprising characteristics of cassava is its wide range of moisture tolerance from high rainfall in the Amazon to regions with a large water deficit such as the semi-arid areas of the Brazilian Northeast. In addition to the consumption of roots, industrial demand for cassava starch is very high (Jansson *et al*. 2009) for use in the food, textile, pharmaceutical, paper, cosmetics, building materials and ethanol industries (Tonukari 2004; Radhika *et al*. 2014).

Cassava, like other crops, is affected by pests and diseases caused by bacteria, viruses, fungi, phytoplasmas and oomycetes (Soto *et al*. 2015). In Brazil, several diseases affect crop yield, such as cassava root rot (CRR), caused by a complex of soil pathogens (Bandyopadhyay *et al*. 2006; Alleyne *et al*. 2015; Vilas Boas *et al*. 2016), that can affect up to 100% of production. Among the most common soil pathogens related to root rot are: *Fusarium oxysporum*, *Pythium*

* To whom all correspondence should be addressed. Email: eder.oliveira@embrapa.br

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press

sp., *Rhizoctonia solani*, *Streptomyces saran* and *Scytalidium lignicola* (Mazzola & Gu 2002; Silva *et al*. 2013; Barros *et al*. 2014). In Brazil, the most prevalent fungi are *Phytophthora* spp., commonly associated with cassava soft rot, and *Fusarium* spp., associated with cassava dry rot root, compromising 30–70% of production losses, respectively (Barros *et al*. 2014).

The search for resistant cassava varieties is one of the major objectives of breeding programmes (Bi *et al*. 2010; Kaweesi *et al*. 2014; Carmo *et al*. 2015; Hokanson *et al*. 2016; Vilas Boas *et al*. 2016) because it is a viable and low-cost alternative to chemical control, especially for low-income farmers. Chemical control is expensive and often inefficient, particularly for soil pathogens, as is the case of root rot disease. On the other hand, recent results indicate progress in the identification of sources of resistance to CRR (Vilas Boas *et al*. 2016).

A number of different breeding strategies have been proposed for the introduction of resistance in cassava in recent decades: such as the evaluation and selection of local varieties, introduction of germplasm, intra- and inter-specific hybridization, use of double haploids, genetic transformation and mutation, the main method of improvement used in cassava involves the crossing of contrasting genotypes and selection of the best individuals through clonal propagation (Ceballos *et al*. 2012, 2015; Oliveira *et al*. 2012). However, it is necessary to use new breeding strategies to ensure higher and continuous genetic gains over succeeding breeding generations.

Molecular polymorphisms have been used to support conventional breeding, aimed at increasing the accuracy and precocity for selecting individuals of interest in order to minimize extensive phenotypic evaluations, which are usually costly and time-consuming (Mengistu *et al*. 2016). In cassava, molecular polymorphisms have been used to diagnose diseases, for analysis of genetic diversity, for determination of post-harvest physiological deterioration, for molecular-assisted selection and for genomic selection (Oliveira *et al*. 2012, 2014; Ceballos *et al*. 2015; Ferguson *et al*. 2015; Wolfe *et al*. 2016). From the molecular point of view, these studies have been increasingly accessible to diverse crops due to the important revolution in the development of high-performance, low-cost genotyping systems. Among these markers, single-nucleotide polymorphisms (SNPs) have become extremely popular in plant molecular genetic studies due to their abundance in the genome and automation in the genotyping process (Mammadov *et al*. 2012).

Although there have been important methodological advances in the use of molecular markers in cassava breeding, little progress has been made in the association between the molecular polymorphism and the phenotypic resistance of germplasm to CRR disease. In contrast, associative mapping or the genome-wide association study (GWAS) has become the main method to identify genes associated with agronomic traits, especially those of complex nature, such as resistance to diseases (Iquira *et al*. 2015). In these studies, molecular polymorphisms provide information for the detection of genomic regions linked to quantitative traits and further allow for the exploration of natural or induced genetic variations (Elshire *et al*. 2011; Poland & Rife 2012; Dell'Acqua *et al*. 2014; Rabbi *et al*. 2014).

Genome-wide association study is based on linkage disequilibrium (LD) and has become a promising strategy to identify genes associated with soil diseases in species such as soybean (Li *et al*. 2016; Zhang *et al*. 2016), maize (Zila *et al*. 2014), wheat (Zegeye *et al*. 2014; Zanke *et al*. 2015; Arruda *et al*. 2016), sugarcane (Singh *et al*. 2016) and beans (Wu *et al*. 2015). It draws the most from historical recombination events to identify the genetic loci that control complex traits using high-density markers and that allow for the selection of parents to create segregant populations and the selection of highly significant SNPs for use as fixed effects on genomic selection models (Begum *et al*. 2015).

Although the genetic diversity of cassava shows the wide natural phenotypic diversity that can be used in the genetic improvement of this species, progress in molecular breeding has been limited mainly to insufficient genomic studies (Rabbi *et al*. 2015). Therefore, it is necessary to identify genomic regions associated with traits of greater interest in order to contribute to the selection of elite plants in breeding programmes. Therefore, the current work aimed to investigate the effect of molecular polymorphisms in resistance of cassava to root rot disease using GWAS. The evaluations were carried out in a set of diversified cassava accessions to provide a comprehensive view of the cassava population structure and its association with some agronomic traits in cassava. Additionally, genetic diversity, dynamics of LD and their effects on GWAS associated with resistance to root rot disease in cassava were analysed.

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press

MATERIAL AND METHODS

Plant material

A total of 263 accessions were analysed, belonging to the Cassava Germplasm Bank (CGB) of Embrapa Mandioca e Fruticultura (Cruz da Almas – Bahia – Brazil), from Brazil (22 different states), as well as Colombia, Venezuela, Nigeria and Thailand. These accessions, composed of landraces and improved varieties, were genotyped and used in the evaluation of resistance to the damage caused by dry root rot (DRR), soft root rot (SRR) and black root rot (BRR) diseases in cassava (Supplementary Table S1).

Experimental design and phenotypic data

The roots used in the evaluation of resistance to damage caused by pathogens in root rot disease were from field trials that followed the cultural practices usually applied to cassava (Souza *et al*. 2006). The planting was carried out at the experimental station of Embrapa Mandioca e Fruticultura (Cruz da Almas – Bahia – Brazil) during the rainy season (March to August). The spacing used was 0·90 m between rows and 0·80 m between plants. Each plot constituted two rows with eight plants, totalling 16 plants per plot (Souza *et al*. 2006; Freitas *et al*. 2016; Vilas Boas *et al*. 2016). Assessments were done in the crop years of 2011, 2012 and 2013 (first and second halves (semesters) of 2011 and 2012 and second semester of 2013). These five evaluation periods were considered independent environments and a minimum of three environments was considered for each accession. The incomplete block design was used in five environments with one observation per plot. Planting was carried out on well-drained soils with no history of root rot disease.

Evaluations for root rot disease were carried out in 2011 and 2012 (first and second semesters) and 2013 (second semester). These five evaluation periods were considered independent environments and the number of evaluated genotypes varied within the seasons. The harvest was carried out manually between 10 and 12 months after planting and, for all trials, there was standardization regarding the size and shape of the roots (20–25 cm in length and 4–6 cm in diameter), respecting the intrinsic characteristics of each germplasm accession.

The isolates used in the inoculations were divided according to the pathogen group and symptoms produced in cassava roots, according to: DRR (*Fusarium solani*, *F. oxysporum* and *Fusarium lateritium*), SRR (*Phytophthora dreschsleri* and *Phytophthora melonis*) and BRR (Family Botryosphaeriaceae – *Neoscytalidium hyalinum*, *Lasiodiplodia theobromae* and *Lasiodiplodia euphorbicola*). A mix of the most aggressive isolates from each group of pathogens was used in the inoculations. The pathogen aggressiveness was previously identified by assessing disease severity in different cassava varieties using several DRR, SRR and BRR species, according to Vilas Boas *et al*. (2016).

To identify sources of resistance to CRR disease, whole roots were washed and disinfected with sodium hypochlorite solution (1%). Subsequently, small holes (approximately 3–4 mm deep) were made in the middle of the roots. The fungi were grown on potato-dextrose agar and oomycetes in V8 medium modified with *β*-Sitosterol. After that, the roots were inoculated at the point of perforation with 100 µl of spore suspension adjusted to 2×10^5 macroconidia/ml. A simulated inoculation (water only) was also used as a control, based on the methodology described by Onyeka *et al*. (2005). After inoculation, the roots were placed in trays containing wet cotton, wrapped in plastic bags (wet chamber) and stored at 26 ± 2·0 °C for 10 days. The experimental design was in randomized blocks with 6–18 replicates per accession (plots represented by one root with three inoculation points, enough to simulate field conditions).

At the time of evaluation, the roots were cut longitudinally and digitally photographed. ImageTool v.3·0 software (UTASCSA, University of Texas Health Science Center, San Antonio, TX, USA) was used as a tool to analyse the symptoms of root rot disease. To calculate the area colonized by the pathogen, two measures were taken: the first regarding the severity of root rot in the peel and the second for severity in the pulp.

The genetic values were predicted by mixed-model analysis and best linear unbiased prediction (BLUP) (Resende 2002) and then used in GWAS. The model $y = Xr + Tb + Zg + Wi + e$ was used for the evaluation of the accessions in an incomplete block design in five environments with one observation per plot, where y is the vector of observations, r is the vector of the environments' effects (random) added to the overall average, b is the block effect within the environment (random), g is the vector of genotypic effects (random), i is the vector of genotype–environment (G × E) effects (random), and e is the vector of residues (random). X, T, Z and W represent the matrix of incidence of the environment, block, accessions and

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press

G × E genotype effects, respectively. The empirical BLUP predictors of genotypic values free of the interaction were given by $\hat{\mu} + \hat{g}_i$ where $\hat{\mu}$ is the overall means of all environments, and $\hat{g}_i$ is genotypic effect free of the G × E interaction.

Estimates of variance components were obtained via restricted maximum likelihood, and individual broad-sense heritability (h_g^2) was obtained according to $h_g^2 = \hat{\sigma}_g^2 / \hat{\sigma}_g^2 + \hat{\sigma}_p^2 + \hat{\sigma}_e^2$, where $\hat{\sigma}_g^2$ = genotypic variance; $\hat{\sigma}_p^2$ = phenotypic variance and $\hat{\sigma}_e^2$ = environment variance.

The package *Corrgram* in R software (R Core Team 2015) was used to estimate the correlations between BLUPs of the severity of root rot diseases in the peel and pulp according to the Pearson coefficient. Correlations between root rot diseases were shown according to similar patterns of disease severity, based on the complexity of resistance and the mechanism of action of the pathogens, which can act jointly or independently.

Genotypic data

The DNA was extracted from leaves of the 263 cassava accessions according to the protocol using cetyltrimethylammonium bromide described by Doyle & Doyle (1990). The DNA concentration was adjusted to 60 ng/µl, and its quality checked by digesting 250 ng of the 10 DNA samples with the restriction enzyme *Eco*RI (New England Biolabs, Boston, MA, USA) at 65 °C for 2 h. Then the DNA samples were taken to the Genomic Diversity facility at Cornell University, USA (http://www.biotech.cornell.edu/brc/genomic-diversity-facility) for library preparation, sequencing and bioinformatics analysis. The basic genotyping by sequencing (GBS) protocol is described by Elshire *et al.* (2011). In summary, the DNA was digested with an *Ape*KI enzyme, a type II restriction endonuclease that recognizes a degenerate five-base sequence (GC*W*GC, where *W* is A or T), and creates a 5′ overhang (3 bp). The linkage between *Ape*KI-cut genomic DNA and adapter was made after the digestion of the samples and the 192-multiplex samples for sequencing. The GBS was performed using *Genome Analyser 2000* (Illumina, Inc., San Diego, CA, USA).

Genome-wide association study and linkage disequilibrium analysis

A priori, the SNPs were submitted to a quality control where all indels and SNPs with >0·10 missing data and minor allele frequency <0·05 were excluded from the analysis. Then the data were imputed using the *missForest* package (R software; R Core Team 2015). *missForest* adopts a strategy that does not require prior information ordering of markers and is indicated for species that have an incomplete or no reference genome. Briefly, the SNP imputation was performed by regressing each variable against all other variables and then predicting missing data for the dependent variable using the fitted forest. Its accuracy has already been tested and performance was found to be superior to other methods and consistent in different data sets (Rutkoski *et al.* 2013).

Linkage disequilibrium was estimated using the correlation coefficients (r^2) between the pairs of loci in each chromosome, and then the distribution pattern for the whole genome was visualized using graphs generated by the *LDheatmap* package (R software; R Core Team 2015). To investigate the decline of LD, r^2 values were plotted as a function of genetic distance, in base pairs (bp), using non-linear regression according to Weisberg (2005).

The association analyses were performed with 14·094 SNPs using the genome association and prediction integrated tool package (R software; Lipka *et al.* 2012; R Core Team 2015). Based on the principle component analysis analysis, *K* (number of sub-populations) was used to construct the *Q* matrix and as a covariate variable (fixed effect) to account for effects due to population structure. The kinship matrix (*K*) was calculated according to VanRaden (2008) and used in a mixed linear model (MLM) as the variance–covariance matrix between the individuals to improve the statistical power of the GWAS model. Moreover, the optimum compression MLM and P3D options (i.e. genetic variance and residual variance for each marker previously determined; Wang *et al.* 2014) were used to reduce the computational time and speed up the statistical power in GWAS.

The efficient mixed-model association algorithm (Kang *et al.* 2008) was implemented to reduce computational time when estimating the components of variance for each marker in the model. The associations between genotype and phenotype were determined using MLM, considering the following equation:

$$Y = X\beta + Zu + e$$

where: Y is an $n \times 1$ vector of observed phenotypes; β is a $q \times 1$ vector of fixed effects (principal components used for population structure estimation, from SNP

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press

markers); u is the vector of random genetic effects; e is the random vector of residual effects; X is the incidence matrix ($n \times q$) relating the individuals to the fixed marker effects β; Z is an $n \times t$ incidence matrices for u, which is the random effect of the mixed model with $\mathrm{Var}(u) = \sigma^2_k$; where K is the $t \times t$ kinship matrix, and e is an $n \times n$ matrix of residual effect, $\mathrm{Var}(e) = \sigma^2_e I$, and the phenotypic variance–covariance matrix can be represented as $V = \sigma^2_g ZKZ' + \sigma^2_e I$.

The MLM allows the elimination of type I errors, avoiding spurious associations (Zhang *et al.* 2010; Wang *et al.* 2014). Significant associations between the markers and the phenotype were estimated using the $-\log_{10}$ P-value, and the Bonferroni correction was adopted for multiple tests, at $P < 0{\cdot}05$.

In silico annotation of single-nucleotide polymorphisms

The putative biological functions of significant SNPs were determined by searching the sequences of the SNPs against proteins related to disease resistance in the Phytozome database (http://www.phytozome.net), in which the homology search was performed using blastx.

RESULTS

Variation and correlation between the severity of pathogens causing root rot disease

Quite a large variation was seen among the evaluated accessions for BLUPs of CRR severity in the peel and pulp (Table 1). However, the pathogens causing dry rot (*Fusarium* species) were more aggressive in comparison with those causing soft rot (*Phytophthora* species) and black rot (Botryosphaeriaceae family species). Hereafter, the pathogens causing DRR, SRR and BRR diseases will be named Fspp, Pspp and Bspp, respectively. In addition, the severity of these pathogens was higher in the peel compared with the root pulp.

Estimates of individual broad-sense heritability of genotypic effects (h^2_g) were of medium magnitude for all groups of pathogens, with similar estimates for Fspp Peel, Pspp Peel, Pspp Pulp, Bspp Peel and Bspp Pulp, while the estimate for Fspp Pulp was half that of the others (Table 1).

According to the BLUP distribution for resistance to CRR, most of the accessions were highly susceptible (distribution in the right on the graphs of Fig. 1).

Table 1. *Average, minimum and maximum values (mm^2) of the best linear unbiased prediction (BLUP) and individual broad-sense heritability of the genotypic effects (h^2_g) for resistance to cassava root rot disease caused by Fusarium spp. (Fspp), Phytophthora spp. (Pspp) and Botryosphaeriaceae (Bspp) species in the peel and pulp of the roots*

Traits	Minimum	Maximum	Average	h^2_g
Fspp Peel	31·6	450·3	69·3	0·31 ± 0·028
Fspp Pulp	20·1	144·8	40·3	0·16 ± 0·019
Pspp Peel	15·8	177·7	41·3	0·26 ± 0·027
Pspp Pulp	7·8	124·9	24·9	0·30 ± 0·026
Bspp Peel	18·3	140·5	50·1	0·28 ± 0·027
Bspp Pulp	10·7	94·1	37·4	0·27 ± 0·026

Only BRR (Bspp in the pulp) presented a distribution closer to normal. In relation to Pearson's correlation estimates, positive associations were identified among all pairs of diseases considering both peel and root pulp evaluations (Fig. 1). The highest correlations (>0·40) were observed for the same pathogen group but in different root parts; for example, Fspp Peel *v.* Fspp Pulp (0·49) and Pspp Peel *v.* Pspp Pulp (0·41). These data indicate the low correlation between the pathogens causing root rot disease in cassava and that this is more interesting between different parts of the plant, which may indicate a presence of different mechanisms of resistance.

Population structure and linkage disequilibrium of single-nucleotide polymorphisms

The SNP markers were distributed on the 18 chromosomes of *M. esculenta* with an average of 26·2 SNP/Mb. Distribution of the SNPs was not uniform and there was variation within the chromosomes, with a minimum of 20·0 SNP/Mb on chromosome 12 and a maximum of 31·1 SNP/Mb on chromosome 3 (Table 2). In addition, 575 SNPs were identified in genomic fragments not yet mapped in the cassava genome, whose genomic coverage was lower in relation to the other chromosomes.

The clustering of the 263 cassava accessions based on the kinship matrix (VanRaden 2008) was used to represent the population structure. Four major genetic clusters were identified (Fig. 2). However, there was no clear distinction between cassava germplasm considering similar characteristics in terms of root pulp colouration (white, cream and yellow),

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press

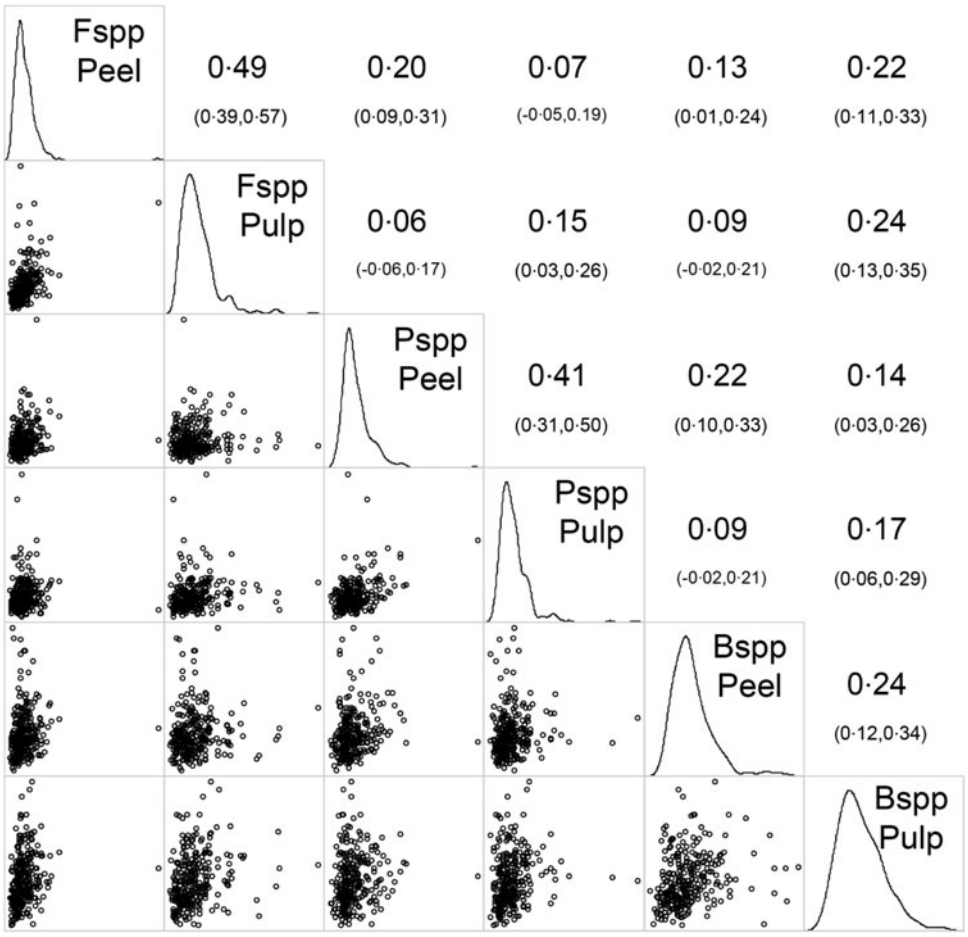


Fig. 1. Correlogram of the best linear unbiased predictor (BLUP) for resistance to cassava root rot disease caused by *Fusarium* spp. (Fspp), *Phytophthora* spp. (Pspp) and family Botryosphaeriaceae (Bspp) in the peel and pulp of the roots. The numbers refer to Pearson coefficients with a 95% confidence interval (in parentheses).

classification of cyanogenic compounds (sweet, intermediate and bitter) and origin of the material (landrace or improvement variety) (Table 3). Considering the genomic kinship matrix, no clear distinction in cassava accessions based on the major or minor susceptibility to root rot pathogens was observed, except for soft rot (Pspp) in the peel, where there is a trend towards the greater susceptibility from group 1 to 4 (Table 3).

Considering only significant r^2 values ($P < 0{\cdot}001$) as informative, it was observed that about 0·04 among all pairs of loci met these restrictions. Significant intra-chromosomal r^2 values ranged from 0·08 to 0·99, with an average of 0·22. Considering only the SNP pairs in significant LD, 0·32 had $r^2 > 0{\cdot}2$, while 0·11 and only 0·02 presented $r^2 > 0{\cdot}5$ and 0·9, respectively. Linkage disequilibrium declined rapidly as a function of genetic distance (base pairs), and, on average, the intra-chromosomal LD reduced with $r^2 = 0{\cdot}2$ near to 15 kb (Fig. 3).

Association analysis

Thirty-eight SNPs were significantly associated with the three classes of CRR evaluated. These SNPs were clustered into 31 loci located on different chromosomes, since SNPs with a difference of up to 1 kb were considered single loci (Table 4). The largest number of significant SNPs (22) was identified for the root rot severity caused by Fspp in the peel. These 22 SNPs were distributed in 17 loci, on chromosomes 2, 5, 6, 8, 9, 10, 12, 17, 18 and 'A' fragments (Fig. 4). However, for the severity of Fspp in the root pulp, eight SNPs were identified in six different loci present in chromosomes 2, 11, 14, 15 and 16. Even with significant SNPs on chromosome 2 for DRR in

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press

Table 2. *Coverage of the single-nucleotide polymorphism (SNP) markers on different cassava chromosomes*

Chromosome	Number of SNPs	Size (Mb)	Genomic coverage (Mb)
1	1121	34·8	32·2
2	910	32·4	28·1
3	1001	29·3	34·1
4	580	28·7	20·2
5	965	28·3	34·1
6	828	27·9	29·7
7	640	27·0	23·7
8	678	33·5	20·3
9	730	29·4	24·8
10	644	26·3	24·5
11	729	27·2	26·8
12	630	31·6	20·0
13	595	28·1	21·2
14	815	24·3	33·6
15	780	26·1	29·9
16	673	29·0	23·2
17	546	27·4	20·0
18	654	25·2	25·9
A	228	32·1	7·1
B	347	31·9	10·9

A and B refer to chromosome fragments not yet mapped in the cassava genome.

the peel and pulp, it was observed that they are placed in different genomic regions. The most significant association for DRR in the peel was observed for SNP C10-20247388 (*P*-value (−log10) = 6·14).

Cassava accessions with the G allele in this SNP presented, on average, a −37·13 mm^2 lesion. For DRR disease in the pulp, the SNP C11-6569478 (*P*-value (−log10) = 4·18) was the most significant, in which accessions with the G allele presented, on average, a −7·06 mm^2 lesion. Considering h_g^2 of 0·31 ± 0·028 and 0·16 ± 0·019 for Fspp in the peel and pulp, respectively, a high number of SNPs significantly associated due to the quantitative nature of these traits are expected. The SNPs explained from 0·05 (13·7 Mb on chromosome 17) to 0·10 (20·2 Mb on chromosome 10) of the phenotypic variation for Fspp in the root peel as well as 0·05 (3·9 Mb on chromosome 2) to 0·06 (6·6 Mb on chromosome 11) of the phenotypic variation for Fspp in the pulp (Table 4).

For Pspp severity in the root peel and pulp, a significant SNP was identified on chromosomes 13 and 8, respectively, which explained 0·08 and 0·07 of the phenotypic variation (Fig. 5). Single-nucleotide polymorphisms C13-28085344 and C08-4077667 demonstrated that accessions having T and G alleles showed, on average, smaller lesions at 11·27 and 6·19 mm^2, respectively (Table 4). For Bspp severity in the root peel, three SNPs were identified on chromosomes 5 and 13 and genomic fragments B, which explained 0·10, 0·09 and 0·10 of the phenotypic variation, respectively. Accessions with G, T and A alleles for the SNPs C05-6181785, C13-2119926 and C20-20622245, respectively, presented lesions of smaller size in relation to the other accessions (−14·4, −12·1 and −8·0 mm^2, respectively) for Bspp severity in the root peel. Three SNPs for Bspp in the root pulp were also identified, associated with two loci present on chromosomes 9 and 12 that explained about 0·05–0·07 of the phenotypic variation of root pulp severity (Fig. 6).

The C, G and T alleles of the SNPs C09-1883587, C09-25235096 and C12-28423630, respectively, were responsible for reducing the size of Bspp lesions in the root pulp at 4·3, 5·2 and 5·3 mm^2, respectively (Table 4). Considering that h_g^2 is of low magnitude and very similar among root rot pathogens (Table 1), it was expected that a greater number of SNPs would be significant for pathogens causing soft and BRR, as occurred with DRR caused by *Fusarium*.

The genetic control of resistance to CRR seems to be quite complex, considering that with the exception of chromosome 19 in the 24·3 Mb region, in which five SNPs closely associated with Fssp in the peel were observed, the other significant SNPs associated with SRR, BRR and DRR were independent and distributed among 14 of the 18 cassava chromosomes. Therefore, the current results suggest that determination of the genetic bases of resistance to root rot caused by these groups of pathogens should be analysed in an individualized way. In addition, the small number of studies on CRR, particularly in the identification of genes or genomic regions associated with resistance, makes it difficult to compare with previously mapped quantitative trait loci (QTL) in the genome of the species.

In silico analysis of single-nucleotide polymorphisms associated with resistance to cassava root rot

Of the 38 SNPs associated with CRR, 19 have biological functions known and available in Phytozome v.11. Specifically for DRR in the peel,

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press

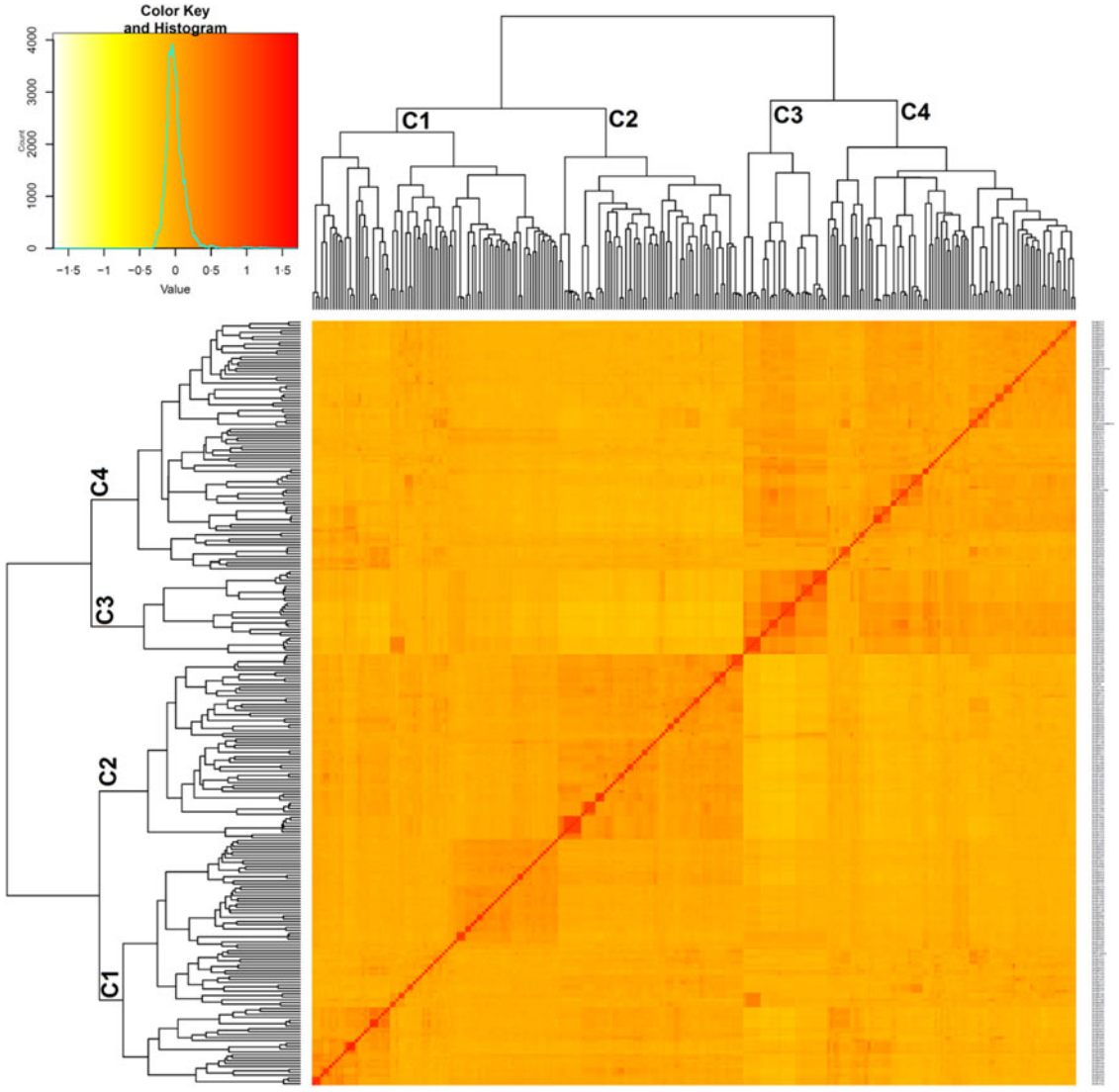


Fig. 2. Relationship between cassava accessions based on the genomic kinship matrix. The four main groups are represented (C1 to C4), having a unimodal distribution of kinship values in different groups. Colour online.

four SNPs associated with proteins without functional annotation were identified, including ten annotations of proteins related to different biological processes such as apoptosis proteins, transcription and translation activation, biosynthetic reactions, cellular signalling cascades, cycle regulation, cells regulating signal transduction, programmed cell death, RING finger domains, phosphatase domain, protein tyrosine kinase, regulation of various physiological programmes including pathogen defence, senescence and trichome development. In the case of DRR in the pulp, three SNPs associated with proteins related to the β-oxidation pathway of fatty acid metabolism, catalytic processes and cellular processes, including division, proliferation, apoptosis and differentiation, were identified (Supplementary Table S2).

For SRR in the peel and pulp, proteins associated with protein recycling from endosomes to the late Golgi and cellular component organization or biogenesis, respectively, were identified. In the case of BRR in the peel, annotations were identified for two SNPs, one of them whose function is unknown and the other associated with the RNA metabolic process, response to endogenous stimulus, response to stress and signal transduction. In contrast, two SNPs associated with BRR in the pulp are related to processes involving translation factors, DNA repair, DNA replication and chromosomal stability (Supplementary Table S2).

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press

Table 3. *Characteristics of the four clusters of cassava accessions based on the kinship matrix obtained with 14·094 single-nucleotide polymorphism (SNP) markers*

Characteristics	Classification	Cluster*			
		1	2	3	4
Pulp colour of roots	White	65·9	93·7	86·2	89·5
	Cream	24·7	4·8	13·8	9·3
	Yellow	9·4	1·6	0·0	1·2
Cyanogenic compound	Sweet	21·2	11·1	34·5	23·3
	Intermediate	7·1	12·7	13·8	17·4
	Bitter	69·4	71·4	51·7	58·1
Origin	Landrace	87·1	93·7	82·8	81·4
	Improvement	12·9	6·4	17·2	18·6
Fspp Peel	Average	66·5	73·7	73·7	66·9
	Range	33·1–144·7	33·7–450·3	36·0–127·5	31·6–145·8
Fspp Pulp	Average	39·6	41·8	38·4	39·8
	Range	20·3–81·7	22·3–144·8	20·1–65·1	21·8–109·2
Pspp Peel	Average	43·1	116·2	163·0	220·5
	Range	20·1–98·7	17·9–81·2	25·6–92·2	15·8–177·7
Pspp Pulp	Average	24·6	24·2	27·4	24·8
	Range	7·8–104·2	8·4–124·9	12·8–59·3	9·9–70·5
Bspp Peel	Average	50·4	46·7	53·1	51·4
	Range	18·3–140·5	21·0–117·7	31·3–84·2	24·2–131·8
Bspp Pulp	Average	36·8	35·9	39·0	38·2
	Range	10·7–85·5	11·1–75·1	20·2–75·3	11·2–94·1

* According to the clustering of Fig. 2.

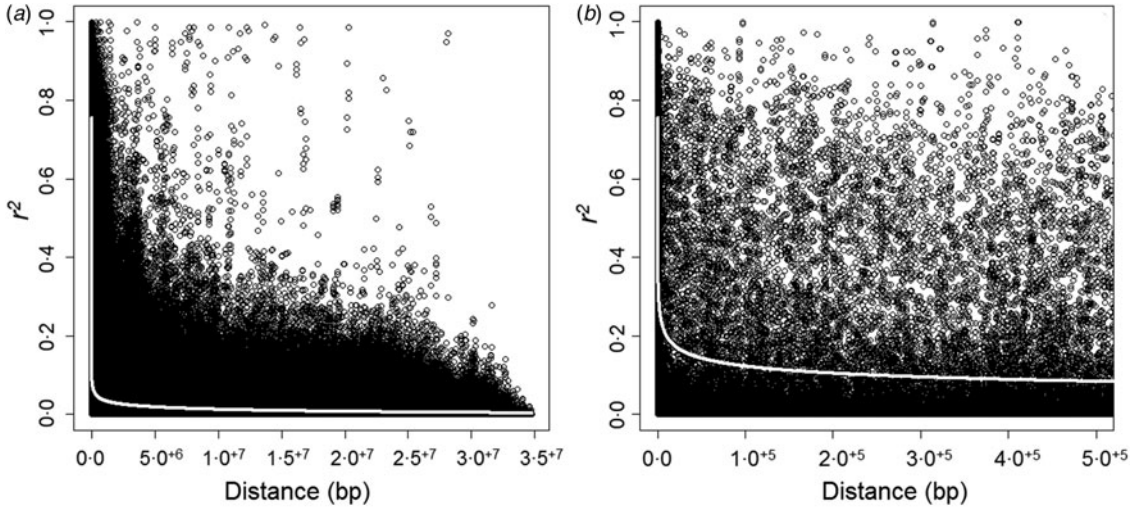


Fig. 3. Decline of linkage disequilibrium (LD– r^2) between SNPs pairs as a function of physical distance in base pairs (bp). (*a*) LD overview on all chromosomes, and (*b*) LD zoom based on distance up to 500 kb. The horizontal line indicates the adjustment of the nonlinear regression for LD along the cassava genome.

DISCUSSION

Severity and genetics of the resistance to cassava root rot

Among the major cassava diseases, root rot has received little research attention; therefore, sources of resistance, pathological and molecular interaction have not been studied much. In several situations, the symptoms of soil diseases are not visible, and thus their impact is not so immediately perceived in

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press

Table 4. *Single-nucleotide polymorphism (SNP) markers associated with cassava root rot disease*

Root rot	SNP	Chromosome	Position	*P*-value (−log10)	Alleles	PEV	Effect
Fspp (Peel)	C02-3825756	2	3825756	4·19**	G/A	5·96	−26·15
	C02-11009956	2	11009956	4·56**	T/C	6·63	−25·35
	C02-17059183	2	17059183	5·40**	T/A	8·17	−34·26
	C05-22175060	5	22175060	3·49*	A/G	4·71	18·44
	C05-26505500	5	26505500	3·74*	T/C	5·15	−15·68
	C06-479302	6	479302	3·85*	T/A	5·35	−25·96
	C06-4555546	6	4555546	3·66*	T/C	5·00	−23·05
	C06-7595885	6	7595885	3·65*	C/T	4·99	22·00
	C06-27475715	6	27475715	4·21**	A/G	5·98	27·39
	C08-6489077	8	6489077	3·74*	G/T	5·15	24·13
	C09-26379964	9	26379964	3·77*	A/C	5·20	22·03
	C10-20247388	10	20247388	6·14**	G/C	9·54	−37·13
	C12-29433550	12	29433550	3·89*	A/G	5·41	19·14
	C12-29433552	12	29433552	3·89*	C/G	5·41	19·14
	C17-13740110	17	13740110	3·46*	C/T	4·66	24·00
	C17-17137567	17	17137567	4·39**	G/A	6·30	−27·23
	C18-1294867	18	1294867	3·55*	G/T	4·81	18·03
	C19-24349376	19	24349376	5·05**	C/G	7·51	−38·32
	C19-24349383	19	24349383	5·04**	A/T	7·51	−38·32
	C19-24349385	19	24349385	5·04**	A/G	7·51	−38·32
	C19-24349391	19	24349391	5·04**	A/G	7·51	−38·32
	C19-24349398	19	24349398	5·04**	A/C	7·51	−38·32
Fspp (Pulp)	C02-3970671	2	3970671	3·48*	A/T	4·87	7·79
	C02-16495400	2	16495400	3·73*	A/T	5·33	6·82
	C11-6569477	11	6569477	3·47*	C/G	4·86	6·22
	C11-6569478	11	6569478	4·18**	C/G	6·13	7·06
	C14-8581966	14	8581966	3·81*	A/C	5·46	10·41
	C14-8582011	14	8582011	3·81*	T/C	5·46	−10·41
	C15-2901287	15	2901287	3·77*	G/A	5·40	7·19
	C16-24353436	16	24353436	3·48*	A/G	4·88	8·08
Pspp (Peel)	C13-28085344	13	28085344	3·75*	C/T	7·99	11·27
Pspp (Pulp)	C08-4077667	8	4077667	3·50*	G/A	6·80	−6·19
Bspp (Peel)	C05-6181785	5	6181785	4·02*	G/A	9·78	−14·43
	C13-2119926	13	2119926	3·49*	C/T	8·90	12·09
	C20-20622245	20	20622245	4·20**	G/A	10·09	8·00
Bspp (Pulp)	C09-1883587	9	1883587	4·01*	T/C	6·30	4·26
	C09-25235096	9	25235096	3·54*	G/A	5·47	−5·18
	C12-28423630	12	28423630	4·42**	T/C	7·04	−5·28

PEV, predictive error variance.

*, **: SNPs significantly associated at $P < 0{\cdot}05$ and $P < 0{\cdot}001$ of probability by the Bonferroni test, respectively. Pathogens causing dry root rot (Fspp), soft root rot (Pspp) and black root rot (Bspp).

comparison to the foliar diseases (Bandyopadhyay *et al*. 2006). Losses in root yield range from 30 to 70% (Barros *et al*. 2014) depending on the pathogen and environmental conditions. Control of soil diseases based on the use of resistant cultivars has been the most economical and reliable approach to controlling CRR, especially when associated with preventive cultural practices (Vilas Boas *et al*. 2016). Therefore, studies related to the search for sources of resistance and genes related to the interaction with cassava are of fundamental importance to guide breeding programmes during the selection process.

The current results showed that the severity of CRR was not similar among the different groups of pathogens (DRR, BRR and SRR), and the greater aggressiveness of *Fusarium* species may be related to better

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press

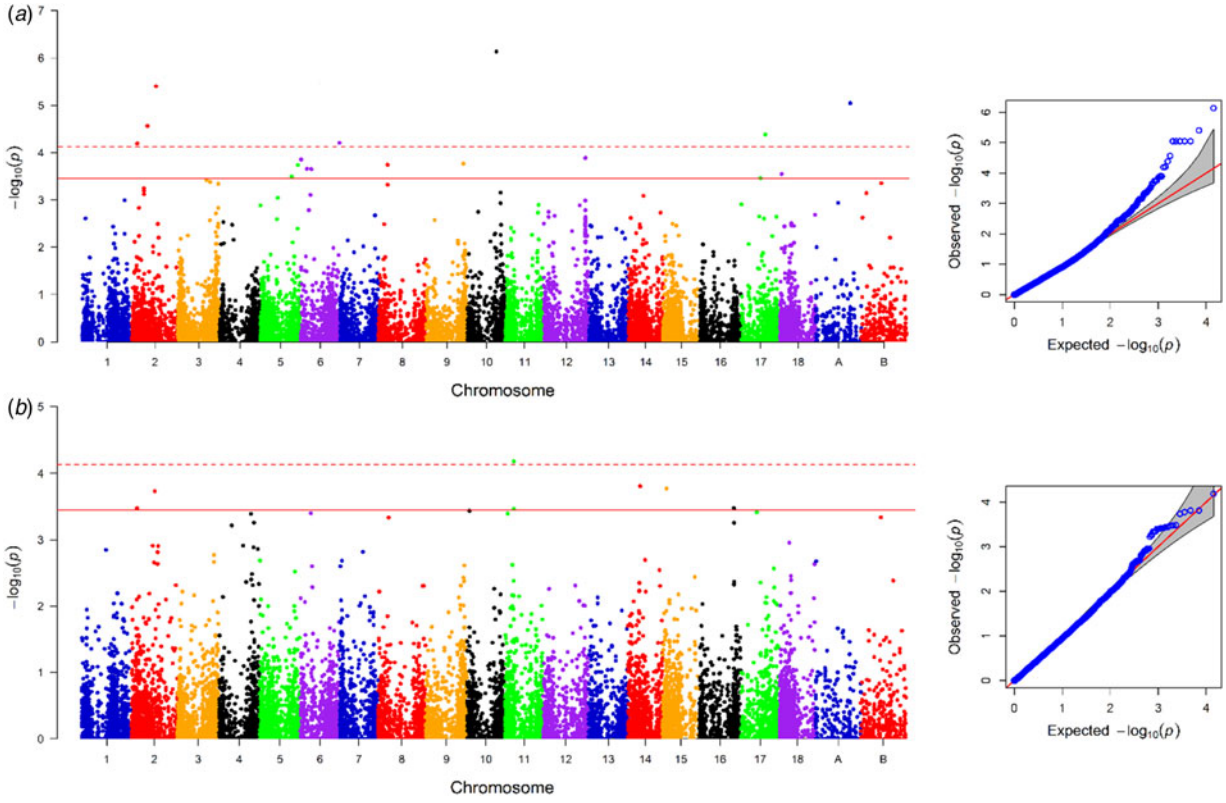


Fig. 4. Manhattan plot indicating SNPs associated with dry root rot (Fspp) in the peel (*a*) and pulp (*b*) of cassava roots. The location of the SNPs on each chromosome (*a* and *b* refer to unmapped SNPs) on the *x*-axis and the association test (−log10 (p)) on the *y*-axis are represented. The dashed line indicates the level of Bonferroni correction at 1% significance and the solid line at 5% significance. The table to the right refers to the quantile–quantile (QQ) plot of the *P*-values observed and expected from the association analysis. Colour online.

adaptation of the pathogen to the host species during infection, compared with the other pathogens, thus providing a greater severity of symptoms. Indeed, Nyaka *et al.* (2015) found that among the main fungi involved in CRR, there was a prevalence of *Fusarium* genus to the detriment of other pathogens (0·50 of fungus sampled).

Differential responses of root rot severity were observed in the different genotypes and also in different parts of the root (peel and pulp). Oliveira *et al.* (2013) also reported the possibility of different levels of resistance among cassava genotypes associated with different levels of *Fusarium* aggressiveness. When studying the prevalence and severity of root rot pathogens in cassava in West Africa, Banito *et al.* (2010) also verified differentiated reactions when the pathogens were inoculated in roots and stems. The virulence tests on cassava stems and roots revealed that *Botryodiplodia theobromae* was extremely virulent when inoculated on stems, whereas *Fusarium* sp., *Sclerotium rolfsii* and *Pythium* sp. were more virulent in the roots. Indeed, Vilas Boas *et al.* (2016) suggested that chemical and physical barriers may be involved in the process of defence against pathogens, resulting in differential responses to the severity of symptoms in different parts of plants.

In relation to h^2_g, the estimates were similar for the different groups of pathogens (0·26–0·31), with the exception of Fspp in the pulp (0·16) (Table 1). These estimates of heritability were of lower magnitude compared to reports for other cassava diseases; for instance, cassava mosaic disease (CMD) resistance. Manu-Aduening *et al.* (2013) evaluated progenies from crosses between 19 landraces varieties of Ghana with a resistant variety (TME11) and reported heritability varying from 0·81 to 0·92 for resistance to CMD. In another study, Parkes *et al.* (2013) reported heritability estimates of median magnitude (0·66) for CMD and bacteriosis in cassava. Therefore, the current results indicate that the nature of cassava resistance to root rot disease is complex and with a strong environmental influence compared with CMD

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press

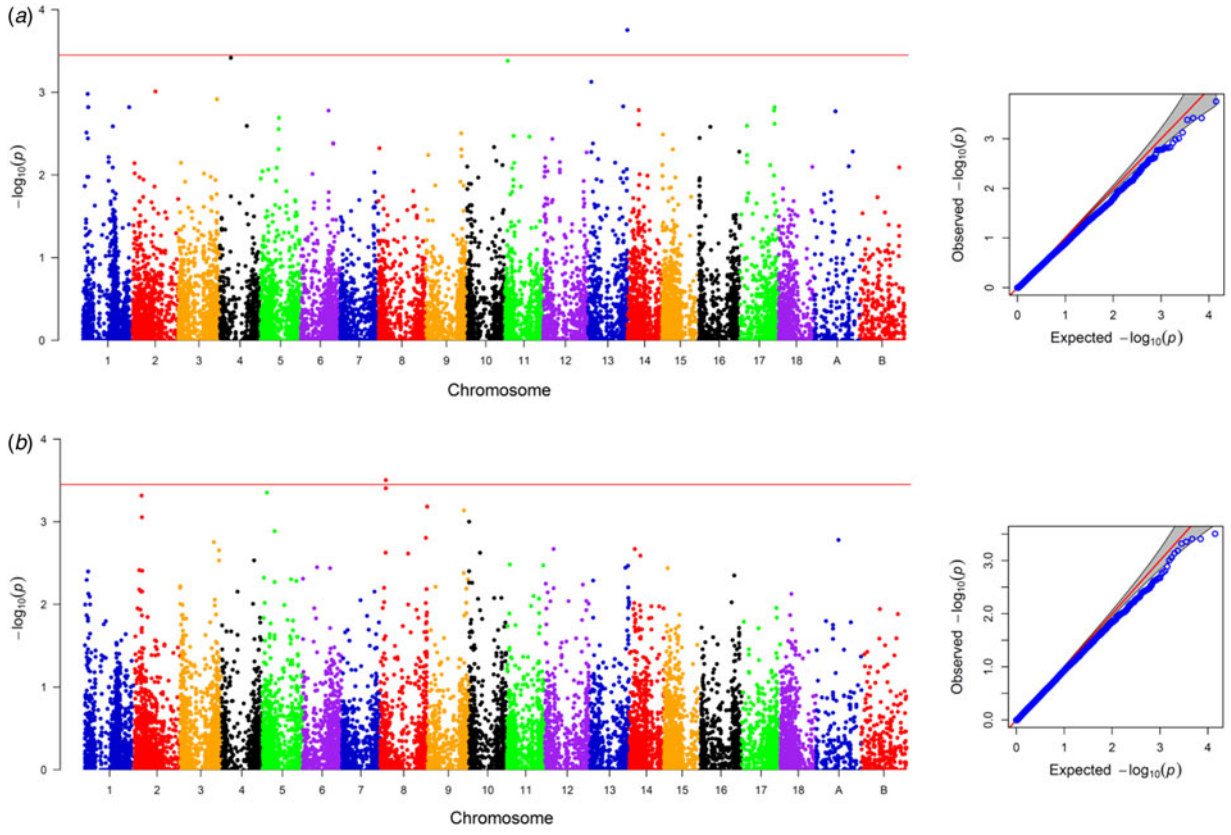


Fig. 5. Manhattan plot indicating SNPs associated with soft root rot (Pspp) in the peel (*a*) and pulp (*b*) of cassava roots. The location of the SNPs on each chromosome (*a* and *b* refer to unmapped SNPs) on the *x*-axis and the association test (−log10 (p)) on the *y*-axis are represented. The solid line indicates the level of Bonferroni correction at $P < 0{\cdot}05$. The table to the right refers to the quantile–quantile (QQ) plot of the *P*-values observed and expected from the association analysis. Colour online.

and bacteriosis. Generally, resistance is associated with the performance of several small-effect genes, making investigation complex, especially when there are several species of pathogens involved, such as in the case of CRR disease. Understanding of the various causal agents associated with complex diseases requires specificity in the understanding of resistance mechanisms (sporulation rate and aggressiveness of each pathogen group) as well as interactions with other pathogens (Bandyopadhyay *et al.* 2006). According to Pariaud *et al.* (2009), many of the components of pathogenic aggressiveness are determined by several genes.

Population structure and linkage disequilibrium

The clusters formed by the kinship matrix show no relationship with agronomic characteristics presented by the cassava accessions, such as the colour of the root pulp, classification of cyanogenic compounds and origin of the material. One reason that may explain the difficulty in clustering cassava accessions based on phenotypic characteristics is the existence of recent and historical hybridizations among accessions belonging to the different clusters that make it difficult to select genotypes with better adaptation to different environments. In fact, because cassava is native to Brazil, the high genetic diversity dispersed in the national territory causes differences in root preference; for example, it is common to find accessions of cassava with yellow pulp and varying levels of cyanogenic compounds.

The same happens with white roots. In addition, during the development of cassava breeding programmes in Brazil, several landraces showing an adaptation worthy of recommendation as a variety for cultivation were collected and evaluated in other production regions. In this case, the landraces have not reached hybridization but only a regionalized introduction and recommendation, so their origin as the improved or local material is not very clear. Regardless of the low association between clustering

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press

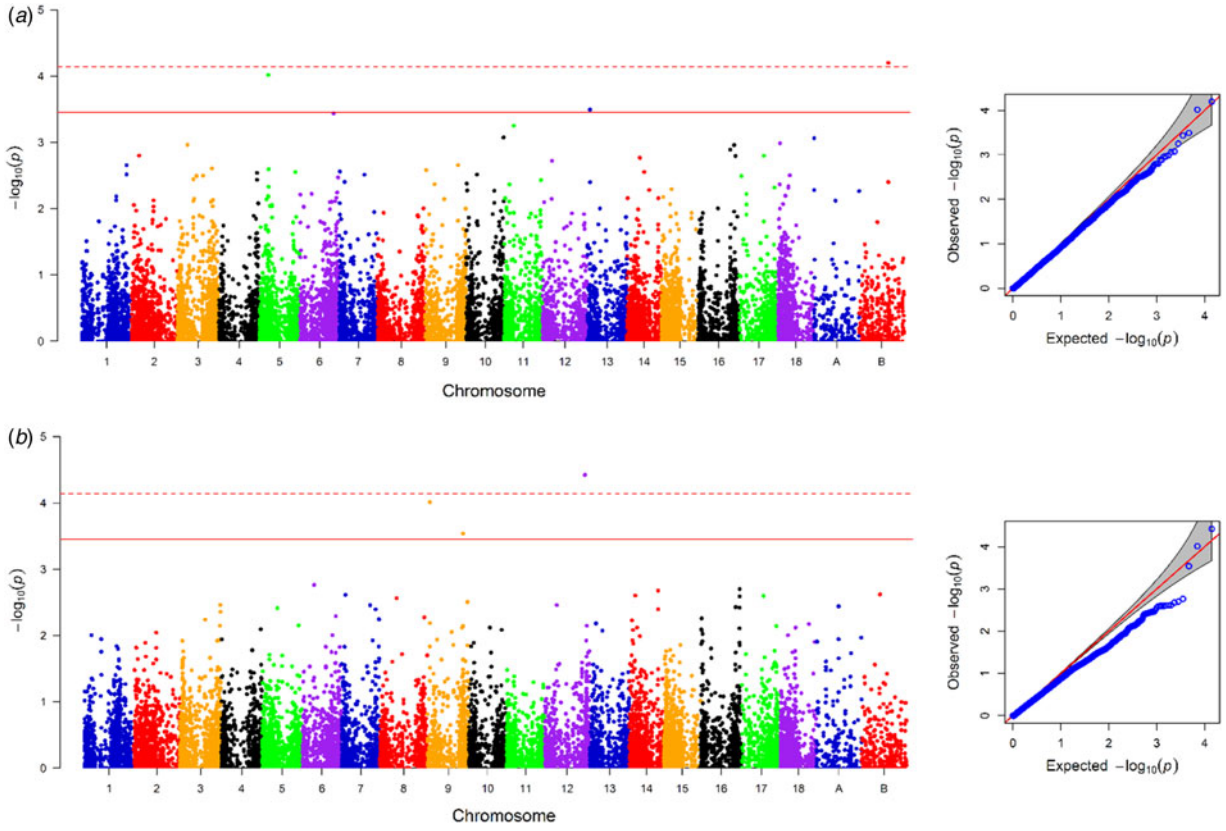


Fig. 6. Manhattan plot indicating SNPs associated with black root rot (Bspp) in the peel (*a*) and pulp (*b*) of cassava roots. The location of the SNPs on each chromosome (*a* and *b* refer to unmapped SNPs) on the *x*-axis and the association test (−log10 (p)) on the *y*-axis are represented. The dashed line indicates the level of Bonferroni correction at 1 $P < 0{\cdot}001$ and the solid line at $P < 0{\cdot}05$. The table to the right refers to the quantile–quantile (QQ) plot of the *P*-values observed and expected from the association analysis. Colour online.

of the genomic kinship matrix with certain agronomic attributes and resistance to root rot disease, the discovery of extensive variation in cassava germplasm due to the formation of four main genetic groups represents an important advance in GWAS in this species and in planning future breeding programmes to capture maximum genetic diversity.

Genome-wide association study basically refers to the identification of QTL based on a statistically significant association between a given phenotype and the molecular marker as a result of LD between the marker and at least one QTL (Gupta *et al.* 2005). Recently, GWAS has been a promising tool in the study of QTLs, since natural populations can be used instead of biparental crosses in order to allow the detection of recombination events in a historical-evolutionary context (Flint-Garcia *et al.* 2005). However, before initiating GWAS, it is important to analyse the distribution of markers, LD and the population genetic structure.

The intra-chromosomal distribution of the SNPs was non-uniformly organized, which probably reflects the history of recombinations that occurred over time in cassava germplasm. In the present study, an average of 700 SNPs was identified by chromosome. Recently, when analysing genomic regions associated with resistance to cassava mosaic geminiviruses, Rabbi *et al.* (2014) found, on average, 256 SNPs by cassava chromosome. Comparison of SNP polymorphism between different studies is difficult, considering that the frequency of SNP detected in the genome seems to be largely dependent on the nature of the genomic region and the individuals studied. In contrast, according to Kawuki *et al.* (2009), in cross-pollinated crops sensitive to inbreeding, such as cassava, high DNA polymorphism is expected, in part due to a large number of loci maintained in the heterozygous state. The high-density and unequal distribution of SNPs contribute to explaining the pattern of LD in

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press

a species, which may affect the accuracy of GWAS (Yan *et al*. 2009).

In addition to the discovery of extensive genetic diversity, LD decay near 15 kb ($r^2 = 0{\cdot}2$) in cassava suggests that GWAS should allow for high-resolution mapping of genes associated with root rot. In other crops, LD decay has been reported to be very close to that observed in the present study, such as date palm (Hazzouri *et al*. 2015) and cranberry (Covarrubias-Pazaran *et al*. 2016), whose estimates are around 10 and 13 kb, respectively. Linkage disequilibrium decay determines the power of QTL detection in associative mapping studies and indicates that the marker density is required to identify significant associations. According to Meuwissen *et al*. (2001), a minimum LD of $r^2 = 0{\cdot}20$ is required for the use of genomic selection to have an accuracy >0·85.

Genome-wide association study

The inheritance of resistance to CRR disease tends to be complex due to the strong environmental influence and possible pathogenic interactions with cassava varieties as well as the differential expression of symptoms in different parts of the roots. Therefore, GWAS is one of the best tools for studying complex traits (Riedelsheimer *et al*. 2012). Accordingly, to investigate the natural genetic diversity of resistance to CRR and to identify genomic regions associated with the disease, high-density GWAS analysis was performed using 14·094 SNPs in a collection of 263 accessions of *M. esculenta*.

The analysis identified 38 significant SNPs, 22 of which are related to the root rot severity caused by Fspp in the root peel, with distribution on nine chromosomes (2, 5, 6, 8, 9, 10, 12, 17 and 18), while, for Fspp severity in the root pulp, eight SNPs were identified in six different loci present on chromosomes 2, 11, 14, 15 and 16. This difference in the number of SNPs associated with *Fusarium* does not seem to be associated with the complexity of the character since the h_g^2 of Fspp in the root pulp is practically half of the Fspp in the root peel (0·16 ± 0·019 and 0·31 ± 0·028, respectively) but presented a smaller number of SNPs associated with the resistance.

The other eight SNPs are associated with Pspp in the peel (1), Pspp in the pulp (1), Bspp in the peel (3) and Bspp in the pulp (3). Other reports have demonstrated that alternative methods of inoculation may result in the presence of different SNPs associated with resistance to the same pathogen. Bonhomme *et al*. (2014) evaluated a collection of 179 *Medicago truncatula* accessions for resistance to *Aphanomyces euteiches* using two complementary phenotyping techniques (*in vitro* and climatic chamber) and identified 44 significant SNPs associated with 32 candidate loci, of which 20 loci were detected only with the *in vitro* technique, seven loci were detected only in the climatic chamber trial and only five were common to both methods. These results indicate that the choice of inoculation method and even of the different parts of the plant to be evaluated can influence the phenotype and consequently result in complex associations at the molecular level. Therefore, specifically in the case of CRR, it will be necessary to select resistant genotypes considering the specificity of the rot groups (SRR, DRR and BRR) and the different parts of the root to guarantee high selective accuracy.

The detection of a higher number of SNPs associated with *Fusarium*, compared with *Phytophthora* and species of the family Botryosphaeriaceae, may indicate a greater complexity of the interaction of this pathogen with cassava. An imbalance in the number of associated markers with disease resistance in other plants has been reported in the literature. For example, Kertho *et al*. (2015) carried out GWAS in wheat to identify resistance to leaf rust caused by *Puccinia triticina* (Pt) and stripe rust caused by *Puccinia striiformis* f. sp. *tritici* (Pst). These authors identified 65 and eight significant markers associated with leaf rust and stripe rust, respectively, being 31 and three QTLs for resistance to Pt and Pst, respectively. However, the use of the panel of higher SNPs density in the cassava genome may contribute to the mapping of other possible regions associated with resistance to *Phytophthora* (SRR) and species of the family Botryosphaeriaceae (BRR). Additional investigations using QTL mapping in biparental (contrasting) populations and further comparative analysis with other inoculation methods and even field-level assessments on soils highly infested by CRR may provide more information on genes and genomic regions that may explain greater part of the phenotypic variation.

Most of the significant SNPs associated with SRR, BRR and DRR were independent and distributed on virtually all cassava chromosomes (14 of the 18 chromosomes). This result seems to be consistent with the uneven distribution of resistance gene analogues (RGAs) in different cassava chromosomes, thus providing evidence of multigene inheritance for the resistance of some cassava diseases (Lozano *et al*. 2015; Soto *et al*. 2015). Recently, Louis & Rey (2015)

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press

demonstrated that tolerance to the South African cassava mosaic virus in the resistant variety (TME3) is associated with a high proportion of RGA transcripts relative to the total differentially expressed genes compared with the susceptible variety (T200).

The present results provide useful information for understanding the genomic structure and mechanisms of action of root rot pathogens in cassava. The SNPs identified in the present study can be used as selection markers for the SNP alleles most associated with the disease in order to increase the resistance levels of the cassava elite varieties.

In silico analysis of single-nucleotide polymorphisms associated with resistance to cassava root rot

Functional annotations of SNPs associated with CRR revealed important genomic regions for positional cloning. From the 19 SNPs related to annotated biological functions, seven of them have functions associated with the plant defence and response to the attack of pathogens (C02-17059183, C02-3825756, C05-6181785, C05-26505500, C06-4555546, C08-6489077 and C18-1294867). Four other SNPs (C06-479302, C06-27475715, C09-26379964 and C16-24353436) relate to protein kinases, an important family of proteins with multiple functions, including disease resistance (Zegeye *et al.* 2014). The other annotations of SNPs associated with CRR are related to general metabolic processes such as RING finger domains; RNA metabolic process, oxidation pathway of fatty acid metabolism, catalysis, cellular component organization, recycling of proteins; transcription regulation, chromosomal stability and DNA repair and replication.

Other authors have reported the involvement of kinases in resistance to diseases in tomato (Andolfo *et al.* 2014), *Solanum* (Zaitsev *et al.* 2002) and wheat (Marone *et al.* 2012). In cassava, Louis & Rey (2015) evaluated the effect of immunity-related genes, including RGA on tolerance and recovery of plants infected with South African cassava mosaic virus; they identified that the contrasting varieties for virus resistance tend to express different RGAs in different sub-cellular targets and that amide-like indole-3-acetic acid and leucine-rich repeat-kinase were the types of resistance protein analogue domains conserved only in the resistant variety (TME3). Therefore, it is very likely that SNPs associated with CRR are associated with protein domains with a function in resistance induction.

In summary, this is one of the first studies on the use of GWAS in understanding the resistance of CRR, in which 38 SNPs associated with the severity caused by DRR (*Fusarium*), SRR (*Phytophthora*) and BRR (species of the family Botryosphaeriaceae) were identified. This research may contribute to further investigations into the existence of causal variants in candidate genes that may affect varietal resistance levels and then provide a basis for cloning of genes related to the resistance to these pathogens in order to understand the complicated molecular mechanism of cassava resistance to root rot disease. However, complementary studies should focus on candidate gene resequencing, fine QTL mapping, complementation studies, panel analysis of cassava germplasm with large numbers of accessions and validation of candidate gene expression in different genetic backgrounds and environments.

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at https://doi.org/10.1017/S0021859617000612.

The authors thank the Fapesb, CAPES and CNPq for the financial assistance and scholarship support.

REFERENCES

Alleyne, A. T., Gilkes, J. M. & Briggs, G. (2015). Early detection of super-elongation disease in *Manihot esculenta* Crantz (cassava) using molecular markers for gibberellic acid biosynthesis. *European Journal of Plant Pathology* **141**, 27–34.

Andolfo, G., Sanseverino, W., Aversano, R., Frusciante, L. & Ercolano, M. R. (2014). Genome-wide identification and analysis of candidate genes for disease resistance in tomato. *Molecular Breeding* **33**, 227–233.

Arruda, M. P., Brown, P., Brown-Guedira, G., Krill, A. M., Thurber, C., Merrill, K. R., Foresman, B. J. & Kolb, F. L. (2016). Genome-wide association mapping of *Fusarium* head blight resistance in wheat (*Triticum aestivum* L.) using genotyping-by-sequencing. *The Plant Genome* **9**, 1–14.

Bandyopadhyay, R., Mwangi, M., Aigbe, S. O. & Leslie, J. F. (2006). *Fusarium* species from the cassava root rot complex in West Africa. *Phytopathology* **96**, 673–676.

Banito, A., Kpémoua, K. E., Bissang, B. & Wydra, K. (2010). Assessment of cassava root and stem rots in ecozones of Togo and evaluation of the pathogen virulence. *Pakistan Journal of Botany* **42**, 2059–2068.

Barros, J. A., Medeiros, E. V., Notaro, K. A., Moraes, W. S., Silva, J. M., Nascimento, T. C. E. S. & Moreira, K. A. (2014). Different cover promote sandy soil suppressiveness to root rot disease of cassava caused by *Fusarium*

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press

solani. *African Journal of Microbiology Research* **8**, 967–973.

Begum, H., Spindel, J. E., Lalusin, A., Borromeo, T., Gregorio, G., Hernandez, J., Virk, P., Collard, B. & McCouch, S. R. (2015). Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (*Oryza sativa*). *PLoS ONE* **10**, e0119873. doi: 10.1371/journal.pone.0119873.

Bi, H., Aileni, M. & Zhang, P. (2010). Evaluation of cassava varieties for cassava mosaic disease resistance jointly by agro-inoculation screening and molecular markers. *African Journal of Plant Science* **4**, 330–338.

Bonhomme, M., André, O., Badis, Y., Ronfort, J., Burgarella, C., Chantret, N., Prosperi, J. M., Briskine, R., Mudge, J., Debelle, F., Navier, H., Miteul, H., Hajri, A., Baranger, A., Tiffin, P., Dumas, B., Pilet-Nayel, M. L., Young, N. D. & Jacquet, C. (2014). High-density genome-wide association mapping implicates an F-box encoding gene in *Medicago truncatula* resistance to *Aphanomyces euteiches*. *New Phytologist* **201**, 1328–1342.

Carmo, C. D., da Silva, M. S., Oliveira, G. A. F. & de Oliveira, E. J. (2015). Molecular-assisted selection for resistance to cassava mosaic disease in *Manihot esculenta* Crantz. *Scientia Agricola* **72**, 520–527.

Ceballos, H., Okogbenin, E., Pérez, J. C., López-Lavalle, L.A.B. & Debouck, D. (2010). Cassava. In *Root and Tuber Crops* (Ed. J. E. Bradshaw), pp. 53-96. New York: Springer.

Ceballos, H., Kulakow, P. & Hershey, C. (2012). Cassava breeding: current status, bottlenecks and the potential of biotechnology tools. *Tropical Plant Biology* **5**, 73–87.

Ceballos, H., Kawuki, R. S., Gracen, V. E., Yencho, G. C. & 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**, 1647–1667.

Covarrubias-Pazaran, G., Diaz-Garcia, L., Schlautman, B., Deutsch, J., Salazar, W., Hernandez-Ochoa, M., Grygleski, E., Steffan, S., Iorizzo, M., Polashock, J., Vorsa, N. & Zalapa, J. (2016). Exploiting genotyping by sequencing to characterize the genomic structure of the American cranberry through high-density linkage mapping. *BMC Genomics* **17**, 451. doi: 10.1186/s12864-016-2802-3.

Dell'Acqua, M., Zuccolo, A., Tuna, M., Gianfranceschi, L. & Pè, M. E. (2014). Targeting environmental adaptation in the monocot model *Brachypodium distachyon*: a multi-faceted approach. *BMC Genomics* **15**, 801. doi: 10.1186/1471-2164-15-801.

Doyle, J. J. & Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus* **12**, 13–15.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, e19379. doi: 10.1371/journal.pone.0019379.

Ferguson, M., Koga, T. M., Johnson, D. A., Koga, K. A., Hirsch, G. N., Becerra López-Lavalle, L. A. & Messier, W. (2015). Identification of genes that have undergone adaptive evolution in cassava (*Manihot esculenta*) and that may confer resistance to cassava brown streak disease. *African Journal of Biotechnology* **14**, 96–107.

Flint-Garcia, S. A., Thuillet, A. C., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., Doebley, J., Kresovich, S., Goodman, M. M. & Buckler, E. S. (2005). Maize association population: a high-resolution platform for quantitative trait locus dissection. *The Plant Journal* **44**, 1054–1064.

Food and Agriculture Organization (FAO) (2013). *Save and Grow: Cassava. A Guide to Sustainable Production Intensification*. Rome, Italy: FAO. Available from: http://www.fao.org/docrep/018/i3278e/i3278e.pdf (accessed 9 February 2017).

Freitas, J. P. X., Santos, V. S. & Oliveira, E. J. (2016). Inbreeding depression in cassava for productive traits. *Euphytica* **209**, 137–145.

Gupta, P. K., Rustgi, S. & Kulwal, P. L. (2005). Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Molecular Biology* **57**, 461–485.

Hazzouri, K. M., Flowers, J. M., Visser, H. J., Khierallah, H. S. M., Rosas, U., Pham, G. M., Meyer, R. S., Johansen, C. K., Fresquez, Z. A., Masmoudi, K., Haider, N., El Kadri, N., Idaghdour, Y., Malek, J. A., Thirkhill, D., Markhand, G. S., Krueger, R. R., Zaid, A. & Purugganan, M. D. (2015). Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. *Nature Communications* **6**, 8824. doi: 10.1038/ncomms9824.

Hokanson, K. E., Ellstrand, N. C., Dixon, A. G., Kulembeka, H. P., Olsen, K. M. & Raybould, A. (2016). Risk assessment of gene flow from genetically engineered virus resistant cassava to wild relatives in Africa: an expert panel report. *Transgenic Research* **25**, 71–81.

Iquira, E., Humira, S. & François, B. (2015). Association mapping of QTLs for sclerotinia stem rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. *BMC Plant Biology* **15**, 5. doi: 10.1186/s12870-014-0408-y.

Jansson, C., Westerbergh, A., Zhang, J., Hu, X. & Sun, C. (2009). Cassava, a potential biofuel crop in (the) People's Republic of China. *Applied Energy* **86** (Suppl. 1), S95–S99.

Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J. & Eskin, E. (2008). Efficient control of population structure in model organism association mapping. *Genetics* **178**, 1709–1723.

Kaweesi, T., Kawuki, R., Kyaligonza, V., Baguma, Y., Tusiime, G. & Ferguson, M. E. (2014). Field evaluation of selected cassava genotypes for cassava brown streak disease based on symptom expression and virus load. *Virology Journal* **11**, 216. doi: 10.1186/s12985-014-0216-x.

Kawuki, R. S., Ferguson, M., Labuschagne, M., Herselman, L. & Kim, D. J. (2009). Identification, characterisation and application of single nucleotide polymorphisms for diversity assessment in cassava (*Manihot esculenta* Crantz). *Molecular Breeding* **23**, 669–684.

Kertho, A., Mamidi, S., Bonman, J. M., McClean, P. E. & Acevedo, M. (2015). Genome-wide association mapping for resistance to leaf and stripe rust in winter-habit hexaploid wheat landraces. *PLoS ONE* **10**, e0129580. doi: 10.1371/journal.pone.0129580.

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press

Khumaida, N., Ardie, S. W., Dianasari, M. & Syukur, M. (2015). Cassava (*Manihot esculenta* Crantz.) improvement through gamma irradiation. *Procedia Food Science* **3**, 27–34.

Li, L., Guo, N., Niu, J., Wang, Z., Cui, X., Sun, J., Zhao, T. & Xing, H. (2016). Loci and candidate gene identification for resistance to *Phytophthora sojae* via association analysis in soybean [*Glycine max* (L.) Merr.]. *Molecular Genetics and Genomics* **291**, 1095–1103.

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S. & Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**, 2397–2399.

Louis, B. & Rey, C. (2015). Resistance gene analogs involved in tolerant cassava–geminivirus interaction that shows a recovery phenotype. *Virus Genes* **51**, 393–407.

Lozano, R., Hamblin, M. T., Prochnik, S. & Jannink, J. C. (2015). Identification and distribution of the NBS-LRR gene family in the cassava genome. *BMC Genomics* **16**, 360. doi: 10.1186/s12864-015-1554-9.

Mammadov, J., Aggarwal, R., Buyyarapu, R. & Kumpatla, S. (2012). SNP markers and their impact on plant breeding. *International Journal of Plant Genomics* **2012**, 1–11 Article ID 728398. doi: 10.1155/2012/728398.

Manu-Aduening, J. A., Peprah, B. B. & Agyeman, A. (2013). Genetic variability of cassava progenies developed through introgression of cassava mosaic disease resistance into Ghanaian landraces. *Journal of Crop Science and Biotechnology* **16**, 23–28.

Marone, D., Panio, G., Ficco, D. B. M., Russo, M. A., De Vita, P., Papa, R., Rubiales, D., Cattivelli, L. & Mastrangelo, A. M. (2012). Characterization of wheat DArT markers: genetic and functional features. *Molecular Genetics and Genomics* **287**, 741–753.

Mazzola, M. & Gu, Y. H. (2002). Wheat genotype-specific induction of soil microbial communities suppressive to disease incited by *Rhizoctonia solani* anastomosis group (AG)-5 and AG-8. *Phytopathology* **92**, 1300–1307.

Mengistu, D. K., Kidane, Y. G., Catellani, M., Frascaroli, E., Fadda, C., Pè, M. E. & Dell'Acqua, M. (2016). High-density molecular characterization and association mapping in Ethiopian durum wheat landraces reveals high diversity and potential for wheat breeding. *Plant Biotechnology Journal* **14**, 1800–1812.

Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829.

Nwokoro, S. O., Orheruata, A. M. & Ordiah, P. I. (2002). Replacement of maize with cassava sievates in cockerel starter diets: effect on performance and carcass characteristics. *Tropical Animal Health and Production* **34**, 163–167.

Nyaka, N. G. A. I. C., Kammegne, D. P., Ntsomboh, N. G., Mbenoun, M., Zok, S. & Fontem, D. (2015). Isolation and identification of some pathogenic fungi associated with cassava (*Manihot esculenta* Crantz) root rot disease in Cameroon. *African Journal of Agricultural Research* **10**, 4538–4542.

Oliveira, E. J., Resende, M. D. V., Santos, V. S., Ferreira, C. S., Oliveira, G. A. F., Silva, M. S. & Aguilar-Vildoso, C. I. (2012). Genome-wide selection in cassava. *Euphytica* **187**, 263–276.

Oliveira, E. J., Ferreira, C. F., Santos, V. S. & Oliveira, G. A. (2014). Development of a cassava core collection based on single nucleotide polymorphism markers. *Genetics and Molecular Research* **13**, 6472–6485.

Oliveira, S. A. S., Hohenfeld, C. S., Santos, V. S., Haddad, F. & de Oliveira, E. J. (2013). Resistance to *Fusarium* dry root rot disease in cassava accessions. *Pesquisa Agropecuária Brasileira* **48**, 1414–1417.

Onyeka, T. J., Dixon, A. G. O. & Ekpo, E. J. A. (2005). Identification of levels of resistance to cassava root rot disease (*Botryodiplodia theobromae*) in African landraces and improved germplasm using *in vitro* inoculation method. *Euphytica* **145**, 281–288.

Pariaud, B., Ravigné, V., Halkett, F., Goyeau, H., Carlier, J. & Lannou, C. (2009). Aggressiveness and its role in the adaptation of plant pathogens. *Plant Pathology* **58**, 409–424.

Parkes, E. Y., Fregene, M., Dixon, A., Boakye-Peprah, B. & Labuschagne, M. T. (2013). Combining ability of cassava genotypes for cassava mosaic disease and cassava bacterial blight, yield and its related components in two ecological zones in Ghana. *Euphytica* **194**, 13–24.

Poland, J. A. & Rife, T. W. (2012). Genotyping-by-sequencing for plant breeding and genetics. *The Plant Genome* **5**, 92–102.

Rabbi, I., Hamblin, M., Gedil, M., Kulakow, P., Ferguson, M., Ikpan, A. S., Ly, D. & Jannink, J. L. (2015). Genetic mapping using genotyping-by-sequencing in the clonally propagated cassava. *Crop Science* **54**, 1384–1396.

Rabbi, I. Y., Hamblin, M. T., Kumar, P. L., Gedil, M. A., Ikpan, A. S., Jannink, J. L. & Kulakow, P. A. (2014). High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. *Virus Research* **186**, 87–96.

Radhika, N. K., Sheela, M. N., Devi, A. A., Sreekumar, J., Kumar, T. M. & Chakrabarti, S. K. (2014). Genetic modification for designer starch from cassava. *Journal of Tropical Agriculture* **52**, 1–6.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Resende, M. D. V. (2002). *Genética Biométrica e Estatística no Melhoramento de Plantas Perenes*. Colombo: Embrapa Florestas: Embrapa Informação Tecnológica, Brasília: Embrapa Informação Tecnológica.

Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., Lisec, J., Technow, F., Sulpice, R., Altmann, T., Stitt, M., Willmitzer, L. & Melchinger, A. E. (2012). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature Genetics* **44**, 217–220.

Rutkoski, J. E., Poland, J., Jannink, J.-L. & Sorrells, M. E. (2013). Imputation of unordered markers and the impact on genomic selection accuracy. *G3: Genes, Genomes, Genetics* **3**, 427–439.

Silva, C. A. D., Medeiros, E. V., Bezerra, C. B., Silva, W. M., Barros, J. A. & Santos, U. J. (2013). Interferência da incorporação de matéria orgânica no solo no controle da podridão negra da mandioca, causada por *Scytalidium lignicola*. *Bioscience Journal* **29**, 1823–1831.

Singh, R. K., Banerjee, N., Khan, M. S., Yadav, S., Kumar, S., Duttamajumder, S. K., Lal, R. J., Patel, J. D., Guo, H.,

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press

Zhang, D. & Paterson, A. H. (2016). Identification of putative candidate genes for red rot resistance in sugarcane (*Saccharum* species hybrid) using LD-based association mapping. *Molecular Genetics and Genomics* **291**, 1363–1377.

Soto, J. C., Ortiz, J. F., Perlaza-Jiménez, L., Vásquez, A. X., Lopez-Lavalle, L. A. B., Mathew, B., Léon, J., Bernal, A. J., Ballvora, A. & López, C. E. (2015). A genetic map of cassava (*Manihot esculenta* Crantz) with integrated physical mapping of immunity-related genes. *BMC Genomics* **16**, 190. doi: 10.1186/s12864-015-1397-4.

Souza, L. S., Farias, A. R. N., Mattos, P. L. P. & Fukuda, W. M. G. (2006). *Aspectos Socioeconômicos e Agronômicos da Mandioca*. Cruz das Almas-Bahia, Brazil: Embrapa Mandioca e Fruticultura.

Tonukari, N. J. (2004). Cassava and the future of starch. *Electronic Journal of Biotechnology* **7**, 5–8.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**, 4414–4423.

Vilas Boas, S. A., Hohenfeld, C. S., Oliveira, S. A., Santos, V. S. & de Oliveira, E. J. (2016). Sources of resistance to cassava root rot caused by *Fusarium* spp.: a genotypic approach. *Euphytica* **209**, 237–251.

Wang, Q., Tian, F., Pan, Y., Buckler, E. S. & Zhang, Z. (2014). A super powerful method for genome wide association study. *PLoS ONE* **9**, e107684. doi: 10.1371/journal.pone.0107684.

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edn. Hoboken, NJ: John Wiley & Sons.

Wolfe, M. D., Rabbi, I. Y., Egesi, C., Hamblin, M., Kawuki, R., Kulakow, P., Lozano, R., Carpio, D. P., Ramu, P. & 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**, 1–13. doi: 10.3835/plantgenome2015.11.0118.

Wu, X., Wu, X., Xu, P., Wang, B., Lu, Z. & Li, G. (2015). Association mapping for *Fusarium* wilt resistance in Chinese asparagus bean germplasm. *The Plant Genome* **8**, 1–6. doi: 10.3835/plantgenome2014.11.008.

Yan, J., Shah, T., Warburton, M. L., Buckler, E. S., McMullen, M. D. & Crouch, J. (2009). Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS ONE* **4**, e8451. doi: 10.1371/journal.pone.0008451.

Zaitsev, V. S., Naroditsky, B. S. & Khavkin, E. E. (2002). Homologs of the genes for receptor-like kinase proteins conferring plant resistance to pathogens: resistance gene homologs in *Solanum* species. *Russian Journal of Plant Physiology* **49**, 810–816.

Zanke, C. D., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., Argillier, O., Stiewe, G., Hinze, M., Neumann, F., Eichhorn, A., Polley, A., Jaenecke, C., Ganal, M. W. & Röder, M. S. (2015). Analysis of main effect QTL for thousand grain weight in European winter wheat (*Triticum aestivum* L.) by genome-wide association mapping. *Frontiers in Plant Science* **6**, 644. doi: 10.3389/fpls.2015.00644.

Zegeye, H., Rasheed, A., Makdis, F., Badebo, A. & Ogbonnaya, F. C. (2014). Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploid wheat. *PLoS ONE* **9**, e105593. doi: 10.1371/journal.pone.0105593.

Zhang, J., Song, Q., Cregan, P. B. & Jiang, G. L. (2016). Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theoretical and Applied Genetics* **129**, 117–130.

Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., Bradbury, P. J., Yu, J., Arnett, D. K., Ordovas, J. M. & Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics* **42**, 355–360.

Zila, C. T., Ogut, F., Romay, M. C., Gardner, C. A., Buckler, E. S. & Holland, J. B. (2014). Genome-wide association study of *Fusarium* ear rot disease in the USA maize inbred line collection. *BMC Plant Biology* **14**, 372. doi: 10.1186/s12870-014-0372-6.

https://doi.org/10.1017/S0021859617000612 Published online by Cambridge University Press