High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding

Ismail Y. Rabbi,†, Martha T. Hamblin,*, P. Lava Kumar, Melaku A. Gedil, Andrew S. Ikpan, Jean-Luc Jannink, Peter A. Kulakow

*International Institute for Tropical Agriculture (IITA), Ibadan, Nigeria
†Department of Plant Breeding and Genetics, Cornell University, Ithaca, NY, USA
‡USDA-ARS, R.W. Holley Center for Agriculture and Health, Ithaca, NY, USA

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Cassava mosaic disease
Breeding
Phenotyping
Monogenic resistance
Genotyping-by-sequencing
QTL

ABSTRACT

Cassava mosaic disease (CMD), caused by different species of cassava mosaic geminiviruses (CMGs), is the most important disease of cassava in Africa and the Indian sub-continent. The cultivated cassava species is protected from CMD by polygenic resistance introgressed from the wild species Manihot glaziovii and a dominant monogenic type of resistance, named CMD2, discovered in African landraces. The ability of the monogenic resistance to confer high levels of resistance in different genetic backgrounds has led recently to its extensive usage in breeding across Africa as well as pre-emptive breeding in Latin America. However, most of the landraces carrying the monogenic resistance are morphologically very similar and come from a geographically restricted area of West Africa, raising the possibility that the diversity of the single-gene resistance could be very limited, or even located at a single locus. Several mapping studies, employing bulk segregant analysis, in different genetic backgrounds have reported additional molecular markers linked to supposedly new resistance genes. However, it is not possible to tell if these are indeed new genes in the absence adequate genetic map framework or allelism tests. To address this important question, a high-density single nucleotide polymorphism (SNP) map of cassava was developed through genotyping-by-sequencing a bi-parental mapping population (N = 180) that segregates for the dominant monogenic resistance to CMD. Virus screening using PCR showed that CMD symptoms and presence of virus were strongly correlated (r = 0.98). Genome-wide scan and high-resolution composite interval mapping using 6756 SNPs uncovered a single locus with large effect (R² = 0.74). Projection of the previously published resistance-linked microsatellite markers showed that they co-occurred in the same chromosomal location surrounding the presently mapped resistance locus. Moreover, their relative distance to the mapped resistance locus correlated with the reported degree of linkage with the resistance phenotype. Cluster analysis of the landraces first shown to have this type of resistance revealed that they are very closely related, if not identical. These findings suggest that there is a single source of monogenic resistance in the crop's gene pool tracing back to a common ancestral clone. In the absence of further resistance diversification, the long-term effectiveness of the single gene resistance is known to be precarious, given the potential to be overcome by CMGs due to their fast-paced evolutionary rate. However, combining the quantitative with the qualitative type of resistance may ensure that this resistance gene continues to offer protection to cassava, a crop that is depended upon by millions of people in Africa against the devastating onslaught of CMGs.

© 2013 The Authors. Published by Elsevier B.V. All rights reserved.
drought and poor soils makes it a food security crop in many parts of sub-Saharan Africa (SSA), on-farm productivity of cassava has remained stagnant for many years due to several production constraints. Cassava mosaic disease (CMD), caused by several species of cassava mosaic geminiviruses (CMGs), is the most economically important constraint to cassava in SSA and the Indian sub-continent (Herrera-Campo et al., 2011). Though significant efforts have been expended on combating this disease, it still causes huge losses to production. The most striking example of the devastating potential of CMD to undermine food security in Africa is the severe pandemic that started as an epidemic in Uganda in the 1990s and led farmers to abandon the crop in many parts of the country (Otien-Nape and Thresh, 1998), and later spread to most countries in East and Central Africa (Legg and Fauguet, 2004). The pandemic is characterized by rapid spread through super-abundant Bemicia tabaci vectors (Legg and Ogwal, 1998) and it is associated with a recombinant strain of the East African cassava mosaic virus – Uganda (EACMV-UG) along with African cassava mosaic virus (ACMV) belonging to the genus Begomovirus, within the Geminiviridae family (Harrison et al., 1997).

Important control measures against CMD include roguing of symptomatic plants, use of virus-free planting materials and deployment of resistant varieties. The first two options are not only labour intensive and difficult to implement but also require continuous and long-term intervention. Use of resistant varieties is the most effective solution in mitigating the negative effect of CMD in farmers’ fields because this approach not only reduces yield losses due to the disease, but also reduces levels of the virus inoculum in the farming system particularly in varieties that suppress virus accumulation.

1.2. Sources of resistance to the disease

Currently deployed resistance against CMD in Africa is of two types: (i) quantitative resistance derived from Manihot glaziovii; and (ii) qualitative resistance conferred by a single resistance gene(s). The quantitative resistance was introgressed into cultivated cassava following an unsuccessful worldwide search for resistant clones in the 1930s (Nichols, 1947; Jennings, 1976). Genetic studies reveal that the polygenic resistance from M. glaziovii is recessive with a heritability of about 60% (Jennings, 1976). The second type of resistance, which is conditioned by a single-gene with a dominant effect, was discovered in the 1980s in landraces from Nigeria and other West African countries (Akano et al., 2002; Fregene et al., 2001). These landraces, which display near-immunity against nearly all species of CMGs, are currently maintained in the IITA germplasm collection referred to as the Tropical Manihot esculenta (TM) series. Diversity studies using molecular markers have previously shown that most of the original landraces bearing this qualitative resistance to CMD are genetically very similar if not identical (Fregene et al., 2000; Lokko et al., 2006). This suggests that the genetic base of this type of resistance in the African cassava genepool may be narrow, or even just a single locus. In contrast, the relative ease with which the highly heritable monogenic resistance can be transferred between germplasm through simple crosses, has resulted in its extensive usage in breeding across Africa as well as pre-emiptive breeding in Latin America (Okogbenin et al., 2007). The long-term stability of this single-gene type of resistance in diverse geographical regions with heterogeneous species and recombinants of CMGs is uncertain given the high evolutionary rate of geminiviruses (Duffy and Holmes, 2009).

Several genetic mapping studies have been conducted to find molecular markers linked to the qualitative resistance in the African germplasm. The first study identified two markers, a microsatellite (SSRSY28) and an RFLP (GY1) that flank a single locus named CMD2 at distances of 9 and 8 cM, respectively (Akano et al., 2002). Subsequent to the discovery of the CMD2 locus, later studies have reported several additional markers that are linked to new resistance genes in other genetic backgrounds, including landraces and improved varieties derived from them (Lokko et al., 2005; Okogbenin et al., 2012). However, nearly all of these studies relied on the bulk segregant analysis (BSA) approach and/or very sparse maps for interval mapping analysis. The BSA approach provides little or no information regarding the chromosomal location of the identified markers, making it difficult to ascertain the number of unique locigenes associated with a trait. When sparse maps are used, the confidence interval surrounding a QTL is usually large, making it difficult to determine the precise QTL location. For example, the CMD2-containing linkage group of Akano et al. (2002) had a total of five markers. Lokko et al. (2005) used a linkage map with just 45 markers, of which only three were in the linkage group containing the resistance locus.

The objective of this study was to provide a comprehensive framework for describing the breadth of the genetic base of the single-gene resistance to CMD in the African cassava germplasm. Firstly, a full-sib mapping population segregating for qualitative resistance to CMD was developed and phenotyped for three growing seasons. The population was genotyped using genotype-by-sequencing (GBS), and a dense genetic linkage map with more than 8000 single nucleotide polymorphisms (SNPs) was constructed. Using this resource, a high-resolution genetic mapping of the CMD resistance locus was carried out. The markers previously reported to be linked to CMD resistance were then projected onto the newly generated genetic map. This revealed their genomic locations, and the spatial relationship between them and the mapped resistance locus from the present study. To confirm the relationship among the CMD resistant landraces, cluster analysis was carried out using genome-wide SNP markers.

2. Materials and methods

2.1. Mapping population development, phenotyping and genotyping

A full-sib mapping population segregating for dominant monogenic resistance to CMD was generated by crossing two non-inbred clones. Both parents are elite lines developed by IITA in Nigeria. The female parent, IITA-TMS-011412, is highly resistant to CMD and also rich in pro-vitamin A. Cloned in 1974, the male parent, IITA-TMS-4(2)1425, is an improved variety from a cross between a landrace from Nigeria (TM109, locally known as Oyarugbafunfun) and variety 58308, a hybrid derived directly from recombination of the M. glaziovii × M. esculenta triple-backcrosses (Jennings, 1976). Variety 58,308 was the main source of the quantitative resistance to CMD and produced some of the first generation Tropical Manihot Selection lines (see the discussion section for more background). IITA-TMS-4(2)1425 shows considerable susceptibility to CMD (Fig. 1).

The 180 F1 seeds produced were germinated in sterilized garden soil and transplanted one month after sowing. At maturity, the seedlings were cloned, regardless whether they were infected or not, and planted at Ibadan, Nigeria (7.40° North latitude, 3.90° East longitude) using a randomized complete block design. Each clone was planted in two replicated plots of five stands per plot with plant spacing of 1 m × 0.5 m for three 12-month long cropping seasons established in 2011, 2012 and 2013. Generation-to-generation propagation through cloning was based on use of 12-cm long stem cuttings in both the infected and non-infected F1s. A local landrace that is highly susceptible to CMD, TME117, was planted as spreader row every fifth plot and as border row surrounding the experimental field to facilitate whitefly-mediated inoculation of the F1 population. The Ibadan site is known for high CMD pressure and
the planting period coincide with high whitefly activity providing high probability of natural exposure of plant population to CMD inoculum. Individual plants were evaluated for CMD symptoms at one and three months after planting using a scale ranging from 1 to 5, with one for symptomless plants while five is given for most severe symptoms (severe mosaic and distortion of leaves).

The entire population was screened for presence of CMGs, particularly for the presence of ACMV and EACMCV, the two predominant species prevailing in West Africa, to confirm virus infection in the infected plants. The third fully expanded leaf from the top was sampled from each of the 5 plants per plot; they were pooled and DNA was extracted and analyzed for ACMV and EACMCV using a multiplex PCR protocol (Alabi et al., 2008).

2.2. Genotyping-by-sequencing

DNA was extracted from 180 F1 individuals and the two parents using a modified Dellaporta method (Dellaporta et al., 1983). Genotyping-by-sequencing as described by Elshire et al. (2011) was carried out at the Institute of Genomic Diversity, Cornell University. Briefly, DNAs from the F1 individuals and parents were digested individually with ApeKI restriction enzyme, which recognizes a five base-pair sequence (GCWGC, where W is either A or T). This enzyme was chosen because of its partial sensitivity to DNA methylation, thus avoiding repetitive elements regions, and frequency of DNA-cuting (Elshire et al., 2011). Two 95-plex GBs sequencing libraries were prepared by ligating the digested DNA to unique nucleotide adapters (barcodes) followed by standard PCR. Sequencing was performed using Illumina HiSeq2000. The sequencing reads from different genotypes were de-convoluted using the barcodes and aligned to the version 4.1 of the cassava reference genome (www.phytozone.org/cassava) by using Burrow Wheelers Alignment tool (Li and Durbin, 2009). SNPs were extracted using the GBS pipeline implemented in TASSEL software (Bradbury et al., 2007), and genotypes were called using a custom R script.

2.3. Data analysis

The pseudo-testcross linkage mapping strategy that is employed in the analysis of full-sib mapping populations requires unambiguous scoring of the parental genotypes at each marker. To ensure this, the parental DNAs were sequenced redundantly four times and their Illumina reads were pooled to increase the number and accuracy of the called SNPs. Following alignment of the reads against the reference genome, the SNPs that segregated in the parents as ab x ab (both parents heterozygous), aa x ab (male parent heterozygous), and ab x aa (female parent heterozygous) were extracted. Prior to linkage analysis, standard quality control was used to filter out SNPs from paralogous sequences (i.e. loci which appear as heterozygous in both parents and all progenies). Also filtered were loci showing significant deviation from expected genotypic frequencies based on chi-square test (threshold for removal: \( P < 0.05 \)) as well as those with missing information in more than 20% of the genotyped individuals in the mapping population.

2.3.1. Mapping of GBS-derived ApeKI SNPs

Genetic linkage maps were constructed using JoinMap version 4.1 (Van Ooijen, 2006). Following calculation of pair-wise recombination frequencies, linkage groups were identified using the logarithm of odds (LOD) score of independence between pairs of loci at a threshold of 10. Due to the large number of markers per linkage group, the maximum-likelihood mapping algorithm implemented in Joinmap 4.1 was used to find the order of the markers in the linkage groups. This method is suitable for dealing with large datasets compared to the regression mapping method (Cheema and Dicks, 2009) and incorporates several numerical methods: simulated annealing for estimating the best map order by minimizing the sum of recombination frequencies in adjacent segments; Gibbs sampling for estimation of multipoint recombination frequency, given the current map order; and spatial sampling of loci to reduce the influence of unknown or dominant genotypes as well as potential errors. Simulated annealing was carried out using a chain length of 30,000 with an acceptance probability threshold of 0.25. Gibbs sampling for maximum likelihood estimation of multipoint recombination frequencies (Jansen et al., 2001) was done using chain length of 50,000 after a burn-in length of 20,000.

2.3.2. Phenotypic data analysis

Because the categorical disease severity scores fitted a bi-modal distribution, for statistical analyses (ANOVA and QTL mapping) the trait was converted to a binary variable (either resistant or susceptible). Individuals with categorical CMD severity score larger than one were classified as Affected; all others were classified as Unaffected. A logistic regression model using generalized linear model was used to estimate the effect of the genotype, replication, environment and genotype-by-environment interaction as follows:

\[
y_{ijkl} = \mu + \beta_1 + R_j + G_k + \beta_1^* G_k + e_{ijkl}
\]

where \( y_{ijkl} \) was the phenotype; \( \mu \) the mean; \( \beta_1 \) the year effect; \( R_j \) the replication effect; \( G_k \) the clone effect; \( \beta_1^* G_k \) is the interaction between clone and year and \( e_{ijkl} \) is the residual. Mixed model was used to obtain best linear unbiased predictors (BLUPs) for each genotype for the combined three-year data. The mixed model was computed using the R package lme4 (Vazquez et al., 2010), considering the effects of the genotypes as random, while replications within environments were regarded as fixed because trials were
carried out for three years. Broad-sense heritability for CMD resistance at one and three months after planting was calculated using the formula

\[ H^2 = \frac{\sigma^2_g}{\sigma^2_g + \sigma^2_e} \]

where \( \sigma^2_g \) and \( \sigma^2_e \) are the variance components for the genotype effect and the residual error, respectively, based on individual plants. Correlations were calculated among disease resistance score BLUPs for three growing seasons (2011, 2012 and 2013).

2.3.3. High-resolution mapping of the CMD resistance locus

Mapping of the CMD resistance locus in the present population was carried out within the BLUPs obtained for each year, and across the combined analysis of the 2011, 2012 and 2013 data. QTL analysis was performed using three complementary approaches. Because of high marker density, single marker–trait association for all 6756 SNPs was carried out. The markers were considered as fixed effects in a linear model implemented in the glm function TASSLE (Bradbury et al., 2007). The genome-wide significance threshold for the F-statistic was determined by the Bonferroni method (Bland and Altman, 1995). Secondly, standard interval mapping (intervals step of 2.5 cM) was carried out using the regression mapping function “scanone” implemented in the R/qtl package (Broman et al., 2003). The genome-wide significance (\( \alpha = 1\% \)) for declaring a significant QTL interval was determined using 1000 permutations. The 95\% Bayesian credible interval for the CMD resistance locus was determined using the function “bayesint” implemented in R/qtl. The proportion of phenotypic variance explained by the resistance locus was obtained by fitting a linear model. Thirdly, QTL analysis using the Composite Interval Mapping (CIM) method was carried out with the number of marker covariates set to three. The number of markers for use as co-factors was determined using the automatic co-factor selection function (stepwiseqtl) implemented in R/qtl. The CIM method enabled a reduction in residual variation and thereby increases the resolution of the QTL location and with it the possibility of detecting any additional genomic regions that underly resistance to CMD. Anchoring CMD-resistance-linked markers in the present high-density genetic map

The high-density SNP map developed in this study was used to anchor published loci associated with qualitative resistance to CMD (Table 1). Primer sequences that flank five of these markers (viz. SSRY28, NS198, SSRY106, NS158 and NS169) were used in BLAST searches of the cassava genome release sequence (www.phytozone.org/cassava). Marker positions were interpolated onto the genetic map on the basis of the scaffolds harbouring them (Table 1).

To obtain a linkage/recombination profile of SNPs along the linkage group that bears the CMD resistance locus, pairwise estimates of linkage disequilibrium (\( r^2; \) Flint-Garcia et al., 2003) were calculated for the SNPs from the entire mapping population using the software package Haploview v. 3.31 (Barrett et al., 2005) and plotted in a matrix form.

2.3.4. Genetic relatedness of the CMD-resistant landraces

In addition to the mapping analysis in the bi-parental population, the genetic relatedness was examined among the TME clones that were originally identified to be sources of the resistance to CMG. A total of 2069 GBS markers from 34 clones, including 29 landraces and five TMS clones with the quantitative resistance (as an out-group), were obtained from a previous study (Ly et al., 2013), and used to perform hierarchical clustering using the Euclidean distance between the genotypes. These distances were used to construct a relationship dendrogram of the clones.

3. Results

3.1. Segregation for resistance to CMD in the mapping population

The frequency distribution of the CMD severity scores in the mapping population revealed a bi-modal pattern with two peaks (Fig. 1): nearly half of the progenies and the female parent (IITA-TMS-011412) were resistant to CMD and showed no symptoms while the remainder of the F1 individuals showed disease symptoms ranging from mild (score 2) to severe (score 5). Resistance to CMD found in the female parent is therefore likely to be a single gene with dominant effect. There was little variation within plots with respect to CMD symptom expression: all stands either showed similar symptoms in infected susceptible plots or no symptoms at all in the resistant plots. The consistency in symptom expression is largely due to the clonal origin from infected cuttings which ensures transmission of viruses across cropping cycles. Moreover, there was very little year-to-year variation in terms of CMD incidences: The disease ratings in the 2011, 2012 and 2013 growing seasons were highly correlated (Pearson’s correlation coefficient, \( r > 0.91 \)). This was reflected in the large broad-sense heritability of the resistance trait as measured at one and three months after planting (\( H^2 \) of 0.89 and 0.93, respectively). Analysis of variance using the logistic regression model showed a highly significant effect for clone (\( p < 2.6 \times 10^{-16} \)), while the other factors such as environment (year), clone \times environment interaction and replication were not significant.

Table 1

<table>
<thead>
<tr>
<th>Marker</th>
<th>Primer sequence</th>
<th>Linkage group</th>
<th>Scaffold (v4.1)</th>
<th>Study</th>
<th>Resistance source</th>
</tr>
</thead>
<tbody>
<tr>
<td>S5214,780931</td>
<td>GBS-SNP</td>
<td>16</td>
<td>scaffold05214</td>
<td>Present study</td>
<td>IITA-TMS-011412</td>
</tr>
<tr>
<td>S5214,30911</td>
<td>GBS-SNP</td>
<td>16</td>
<td>scaffold05214</td>
<td>Rabhi et al. (in press)</td>
<td>IITA-TMS-961089A</td>
</tr>
<tr>
<td>S5214,780931</td>
<td>Rev:TTGACATGCTGTAATTTCTTGAG</td>
<td>16</td>
<td>scaffold05214</td>
<td>Akano et al. (2002), Lokko et al. (2005), Okogbenin et al. (2012)</td>
<td>TME3; TME7; IITA-TMS-972205</td>
</tr>
<tr>
<td>S5214,780931</td>
<td>Rev:GTGCTGCAGAAACTAAAT</td>
<td>16</td>
<td>scaffold06906</td>
<td>Okogbenin et al. (2007)</td>
<td>TME3</td>
</tr>
<tr>
<td>S555158</td>
<td>Fw:GTGCGAAATGGGTAATCTACG</td>
<td>16</td>
<td>scaffold06906</td>
<td>Okogbenin et al. (2007)</td>
<td>TME3</td>
</tr>
<tr>
<td>S555158</td>
<td>Rev:GCTTATCATTAGCAGGGCAGC</td>
<td>16</td>
<td>scaffold06906</td>
<td>Okogbenin et al. (2007)</td>
<td>TME3</td>
</tr>
<tr>
<td>S555158</td>
<td>Rev:GCTTATCATTAGCAGGGCAGC</td>
<td>16</td>
<td>scaffold06906</td>
<td>Okogbenin et al. (2007)</td>
<td>TME3</td>
</tr>
<tr>
<td>SRFI KME-1</td>
<td>Fw:ATGGAGAGAGGATGACACTAGT</td>
<td>16</td>
<td>No match</td>
<td>Okogbenin et al. (2007)</td>
<td>TME3</td>
</tr>
<tr>
<td>SRFI KME-1</td>
<td>Rev:AGAAGAGAGGATGACACTAGT</td>
<td>16</td>
<td>No match</td>
<td>Okogbenin et al. (2007)</td>
<td>TME3</td>
</tr>
<tr>
<td>SRFI KME-1</td>
<td>Rev:GGAAGAGGAGGATGACACTAGT</td>
<td>16</td>
<td>Scaffold04175</td>
<td>Okogbenin et al. (2012)</td>
<td>IITA-TMS-972205</td>
</tr>
<tr>
<td>SRFI KME-1</td>
<td>Rev:GGAAGAGGAGGATGACACTAGT</td>
<td>16</td>
<td>Scaffold07933</td>
<td>Lokko et al. (2005)</td>
<td>TME7</td>
</tr>
</tbody>
</table>
3.2. Screening of the population for cassava mosaic geminiviruses using PCR

The PCR-based screening of the mapping population detected one or more of the CMGs in 87 of 180 plots assayed, while 93 plots were negative (Table 2). Only two species of CMGs, ACMV and EACMCV, were detected in the trial, which is consistent with known CMGs prevalence in West Africa. ACMV was detected in all the 87 virus-positive plots, whereas EACMCV was detected as a coinfection with ACMV in 16 plots (17% incidence). No case of single infection by EACMCV was detected. CMGs were detected in 79 of 81 symptomatic plots indicating a strong positive correlation between the visual scoring of disease and the PCR results. In addition, PCR also detected occurrence of ACMV in 7 asymptomatic plots and ACMV and EACMCV in one plot. Mean severity of CMD symptoms in ACMV infected plants was 3.1 and plants infected with ACMV and EACMCV was 3.2, which suggest apparent lack of synergistic effect in heightening symptom severity in dually infected plants.

3.3. Genotyping of SNP markers and construction of a dense genetic map

In all, 17,682 SNPs were obtained from the SNP calling pipeline. The SNP data were subsequently filtered for markers with more than 20% missing values across the genotyped individuals. Also removed were loci that deviated from the expected genotypic frequencies at Chi-square significance threshold of $P < 0.05$. Linkage analysis was done using 8704 SNPs that passed these two QC filters.

A high-density genetic linkage map was constructed using the Maximum-Likelihood approach implemented in Joinmap 4.1 (Table 3). A total of 6756 SNP markers were mapped across 19 linkage groups with between 115 and 559 SNPs (average = 256). With an average inter-SNP distance of 0.52 cM, this is the densest map developed for cassava so far (Fig. 2). Despite the high density of the GBS-derived SNPs mapped, several regions without markers were observed. Most notable were a single region on linkage group 4 and two regions in linkage group 18. A possible reason for such gaps could be lack of polymorphic markers as a result of identity-by-descent of this region in the two parents.

Most of the 12,977 scaffolds that constitute the current 533 Mb of the version 4.1 cassava genome assembly are fairly small; nearly 95% of them are 200 Kb or smaller. A total of 1093 unique scaffolds were anchored in the present map, and ranged from 19 to 89 in the different linkage groups. Despite their relatively small number, the anchored scaffolds covered a total size of 313.3 Mb, and accounted for 58.7% of the current cassava genome assembly. The complete genetic map developed from this work is available in Supplementary Table 1.

3.4. High-resolution mapping of CMD resistance locus

Based on the qualitative nature of the resistance to CMD in the present mapping population, a single locus was expected to underlie the resistance phenotype. The high-density genetic map developed with 6756 GBS SNPs permitted a genome-wide search

![Genetic map](http://dx.doi.org/10.1016/j.virusres.2013.12.028)

Table 2

<table>
<thead>
<tr>
<th>Disease status</th>
<th>Virus not detected</th>
<th>ACMV only</th>
<th>ACMV + EACMCV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptomatic</td>
<td>1</td>
<td>65</td>
<td>15</td>
</tr>
<tr>
<td>Asymptomatic</td>
<td>92</td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3

<table>
<thead>
<tr>
<th>Linkage group</th>
<th>Number of SNPs</th>
<th>Length (cM)</th>
<th>Average distances (cM)</th>
<th>Number of scaffolds anchored</th>
<th>Cumulative scaffold size (base-pairs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>473</td>
<td>242</td>
<td>0.51</td>
<td>66</td>
<td>30,226,735</td>
</tr>
<tr>
<td>2</td>
<td>344</td>
<td>195</td>
<td>0.57</td>
<td>52</td>
<td>20,822,503</td>
</tr>
<tr>
<td>2.2</td>
<td>366</td>
<td>175</td>
<td>0.48</td>
<td>59</td>
<td>20,298,754</td>
</tr>
<tr>
<td>3</td>
<td>419</td>
<td>168</td>
<td>0.40</td>
<td>55</td>
<td>12,608,655</td>
</tr>
<tr>
<td>4</td>
<td>255</td>
<td>194</td>
<td>0.76</td>
<td>57</td>
<td>13,843,915</td>
</tr>
<tr>
<td>5</td>
<td>543</td>
<td>230</td>
<td>0.42</td>
<td>57</td>
<td>15,175,445</td>
</tr>
<tr>
<td>6</td>
<td>275</td>
<td>182</td>
<td>0.67</td>
<td>73</td>
<td>15,911,788</td>
</tr>
<tr>
<td>7</td>
<td>559</td>
<td>256</td>
<td>0.46</td>
<td>66</td>
<td>21,949,504</td>
</tr>
<tr>
<td>8</td>
<td>454</td>
<td>222</td>
<td>0.49</td>
<td>78</td>
<td>18,054,518</td>
</tr>
<tr>
<td>9</td>
<td>207</td>
<td>72</td>
<td>0.35</td>
<td>19</td>
<td>6,148,875</td>
</tr>
<tr>
<td>10</td>
<td>299</td>
<td>148</td>
<td>0.50</td>
<td>55</td>
<td>15,132,961</td>
</tr>
<tr>
<td>11</td>
<td>304</td>
<td>203</td>
<td>0.67</td>
<td>41</td>
<td>13,297,633</td>
</tr>
<tr>
<td>12</td>
<td>115</td>
<td>99</td>
<td>0.87</td>
<td>21</td>
<td>6,111,492</td>
</tr>
<tr>
<td>13</td>
<td>302</td>
<td>200</td>
<td>0.67</td>
<td>60</td>
<td>17,048,413</td>
</tr>
<tr>
<td>14</td>
<td>543</td>
<td>242</td>
<td>0.45</td>
<td>89</td>
<td>19,282,711</td>
</tr>
<tr>
<td>15</td>
<td>451</td>
<td>225</td>
<td>0.50</td>
<td>49</td>
<td>18,831,996</td>
</tr>
<tr>
<td>16</td>
<td>281</td>
<td>152</td>
<td>0.54</td>
<td>64</td>
<td>16,048,872</td>
</tr>
<tr>
<td>17</td>
<td>275</td>
<td>156</td>
<td>0.57</td>
<td>70</td>
<td>17,372,231</td>
</tr>
<tr>
<td>18</td>
<td>291</td>
<td>154</td>
<td>0.53</td>
<td>58</td>
<td>14,665,500</td>
</tr>
<tr>
<td>Total</td>
<td>6756</td>
<td>256</td>
<td>0.52</td>
<td>1093</td>
<td>313,332,501</td>
</tr>
</tbody>
</table>

* Average values per linkage group.
for the locus underlying the qualitative resistance to CMD. A strong association was detected in linkage group 16 that peaked at around 69.12 cM (Fig. 3). This association decreases on both sides of the peak as a result of increasing recombination between the markers and the underlying resistance gene. Most SNPs showing highly significant association ($P < 10^{-4}$) came from the 1.46 Mb-long scaffold 5214, with marker S5214_780931 at the peak; this marker explained 74% of the disease resistance variance. Additionally, only those SNPs that are informative in the CMD-resistant female parent show the significant associations while those segregating only in the susceptible male parent do not (Fig. 3). The presently mapped resistance locus occurs in the vicinity of the previously mapped CMD2 locus (Akano et al., 2002); marker S5214_780931 is just 623.24 kb away from a microsatellite marker, SSRV28 (between 157,470 bp to 157,616 bp), that was first reported to be closely linked to the CMD2 locus (Akano et al., 2002), indicating that the same gene may account for the observed resistance to CMD in the present mapping population.

The results from the interval-mapping based QTL analysis recapitulated those from single-marker trait associations, and uncovered a single peak with a maximum LOD value of 43 on linkage group 16 (Supplementary Fig. 1). Despite employing the Composite Interval Mapping method, no additional peaks exceeding the significance threshold were detected, confirming that only a single locus conferred the qualitative resistance in the present population. The SNP marker S5214_780931 is flanked by S5214_472282 and S5214_1084049 was the closest to the QTL peak (Table 4). The approximate 95% Bayesian credible interval for the mapped locus spans from 68.8–70.74 cM region along LG16, irrespective of the scoring time (one or three months after planting) or growing season (2011, 2012 or 2013). The profiles of LOD scores along the linkage group 16 were very similar for the disease scores recorded at one and three months after planting as well as the three seasons of data, supporting the high heritability observed for this trait. Comparison of the phenotypes of the F1s against the resistance-linked marker S5214_780931 genotypes showed a small proportion of recombinants that carry the SNP allele linked to resistance but show susceptibility to the disease and vice versa.

3.5. Genomic localization of markers flanking previously mapped qualitative resistance against CMD

A major objective of the present study was to use the high-density SNP genetic map to anchor seven molecular markers previously reported to be linked to single dominant gene resistance to CMD in four other genetic backgrounds (Table 1). The SSR and RFLP makers used in those studies were therefore anchored in the present map via their harbouring scaffolds. These markers came from scaffold 05214 (SSRY28), scaffold 06906 (NS158 and NS169), scaffold 04175 (NS198) and scaffold 07933 (SSRY106) all of which fall within the same genomic region of linkage group 16 of the present map (Fig. 4). The primer sequences for the RFLP marker RME-1 did not identify a suitable match in the reference genome. In a parallel study using another bi-parental mapping population derived from a cross between another improved variety that is nearly immune to CMD (IITA-TMS-961089A) and a susceptible landrace (TME117), another SNP marker was identified from the same scaffold (S5214.30911). It was strongly associated with the disease resistance and explained 60% of phenotypic variation (Rabbi et al., in press). The reported percentage of variation explained by these markers shows a gradient that peaks around scaffold 05214, the region that is likely to contain the concerned resistance gene (Fig. 4). Markers away from this region have been reported to be less linked to the resistance gene by the low percentage of variation that they explain, a trend that agrees with the GWA results, particularly considering the segregating markers from the female parent (Fig. 3).

3.6. Recombination pattern in linkage group 16

To visualize the degree of linkage and recombination pattern between the presently mapped CMD resistance locus and other previously published SSR markers along the linkage group, the chromosome-wide pattern of LD on linkage group 16 was examined (Fig. 4). The resulting haplotype map is useful in providing a framework for interpreting the results of the previous mapping studies, particularly the proportion of phenotypic variation explained by the various microsatellite markers and their relative locations in the present map. Though different parents are used in the present and previous mapping studies, these populations have all

Table 4

<table>
<thead>
<tr>
<th>Trait</th>
<th>SNP</th>
<th>Linkage group</th>
<th>Position (cM)</th>
<th>Peak LOD$^a$</th>
<th>$R^2$</th>
<th>$H^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMD15</td>
<td>S5214_1084049</td>
<td>16</td>
<td>68.88</td>
<td>45.37</td>
<td>0.708</td>
<td>0.892</td>
</tr>
<tr>
<td></td>
<td>S5214_780931</td>
<td>16</td>
<td>70.00</td>
<td>45.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S5214_472282</td>
<td>16</td>
<td>70.74</td>
<td>44.99</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Logarithm of odds score for presence of QTL.

$^b$ Values calculated using the three-year BLUPs; $R^2$ = percentage variation explained by QTL; $H^2$ = Heritability using the three year data.
undergone a single round of meiosis and are thus expected to have a similar extent of linkage disequilibrium across their chromosomes.

The region bearing the CMD resistance locus was characterized by two large haplotype blocks (dark-grey shading between 0 to 32 cM and 36 to 62 cM, respectively). Moderate LD was detected between these blocks (light-grey shading). The first block encompasses scaffold 4175, which harbours microsatellite marker NS198, reported by Okogbenin et al. (2012) to explain 11% of variation in CMD resistance in a bi-parental population. SNPs from this scaffold and those near the CMD2 locus in scaffold 05214 also show moderate LD ($r^2 \sim 0.10$).

The CMD resistance locus occurs in the second LD block. Scaffold 5214 harbours the SNPs that were strongly associated to the resistance as well as the microsatellite marker SSRY28 reported previously (Akano et al., 2002). Though discovered from different genetic backgrounds, the resistance-linked SNPs and SSRY28 explain between 60% and 70% of the disease resistance variance. These markers are just 623.24 kb apart. Another scaffold in this block (07933) which harbours microsatellite marker SSRY106, was reported to explain nearly 40% of variation in CMD resistance (Lokko et al., 2005).

3.7. Genetic relatedness of CMD-resistant landraces using genome-wide SNPs

In addition to genetic mapping of the bi-parental population, cluster analysis was performed with key landraces known to possess strong resistance to CMD (Fig. 5). To estimate the “residual genetic distance” between identical clones resulting from genotyping error, several clones – of which have different names as a result of adoption in different regions – were redundantly genotyped. These pairs include the male parent (IITA-TMS-4(2)1425) and “Kibandameno-white”; IITA-TMS-30572 known as “Migyera” in Uganda; and TME12 (A and B). Most of the CMD-resistant landraces, including those that were first discovered in Nigeria, are clearly very closely related, or even perhaps identical, based on comparison of the distance between them and the residual distance between the redundantly genotyped clones, confirming
previous studies using AFLPs (Fregene et al., 2000). These landraces (TME3, TME4, TME6, TME7, TME11, TME12 and TME14) have a very similar morphological appearance; most prominent of which is red petioles.

4. Discussion

4.1. A historical perspective of development and diffusion of early CMD resistant varieties across of Africa

Breeding for resistance to CMD has been a major goal of cassava improvement programmes in Africa for more than 70 years. Prior to the discovery of the single-gene resistance (Akano et al., 2002), the primary defense against the disease was the polygenic resistance introgressed into cultivated cassava (M. esculenta) from M. glaziovii after three cycles of backcrossing (Nichols, 1947). Examining the breeding history of improved varieties with the quantitative resistance to CMD reveals a very narrow genetic base tracing back to a single derivative of the M. glaziovii × M. esculenta crosses (Jennings, 2003). None of these descendants is immune to infection by CMGs (Nichols, 1947) although some express mild and sometimes transient symptoms as a result of incomplete systemic infection that leads to reversion of symptoms (Jennings, 1976; Fargette et al., 1994), while others are quite susceptible to the disease. These TMS varieties that trace back to M. glaziovii have been widely adopted in Africa and helped revive cassava farming following the devastation of many local landraces in East Africa (Legg and Fauquet 2004). Some of the adopted TMS varieties are IITA-TMS-60142, IITA-TMS-30337 and IITA-TMS-30572. Since the 1990s, IITA has been making crosses to combine the single dominant gene (CMD2) with the multigene M. glaziovii resistance and have produced clones with near immunity to CMD (Legg and Fauquet, 2004).

4.2. Mapping resolution of the CMD2 locus

Genetic mapping of qualitative resistance to CMD in this study uncovered a single locus on linkage group 16 with a large peak LOD (>40). This locus, SS214_780931, explained 74% of the phenotypic variance, and is co-located with the SSRY28 on scaffold5214, indicating the dominant monogenic resistance gene from the female parent is likely to be the CMD2 locus of Akano et al. (2002). The dense genetic map obtained from GBS has enabled a higher level of mapping resolution of the CMD resistance locus. Linkage group 16 has a total of 281 SNPs, and the approximate 95% Bayesian credible interval around the mapped locus is just 1.1 cM. This resolution was not obtainable using the traditional marker from previous mapping efforts.

4.3. All markers linked to qualitative resistance occur in the same chromosome region

The high-density SNP genetic map was used to anchor markers previously reported to be linked to the dominant monogenic resistance to CMD. These markers (Table 1) were interpolated in the present map via the scaffolds that harbour them. All of them (except RME1 and RME4 for which a suitable matches in the current reference genome were not found) came from scaffolds occurring in the same region of linkage group 16. Overall, there is a general congruence between the proportion of genetic variation reported for these microsatellite markers, the marker-trait association profile in linkage group 16 (Fig. 3) and the pattern of linkage disequilibrium (measure as r2) between microsatellite locations and the putative position of the resistance locus in the present population (Fig. 4). This pattern supports the idea that most of the microsatellite markers reported to be linked to varying degrees with CMD resistance are associated with a single resistance locus that is most likely the CMD2 gene. Moreover, the strength of the linkage decreased with increasing distance from the gene location. In a parallel mapping study using another improved variety that is nearly immune to CMD (IITA-TMS-961089A) and a susceptible landrace (TME117), a strong QTL signal was found in scaffold05214 that also explained 60% of resistance variation, implicating the same CMD2 locus (Rabbi et al., in press). Considering the results of the present study and those of Akano et al. (2002), Okogbenin et al. (2012), and Lokko et al. (2005), it is highly likely that there is a single gene, or a cluster of resistance genes (Michelmore and Meyers, 1998).

Indeed, the conservation of QTL positions among the different sources of CMD resistance is not surprising given that the majority of the highly resistant cultivars developed recently in Africa trace to a few landrace accessions from Nigeria that were used over the last 12 or so years of resistance breeding in which the merits of the TMS materials were appreciated. According to the phylogeeny (Fig. 5) it was confirmed that several landraces that were first discovered to be nearly immune to CMD were genetically very similar. These findings agree with a previous study (Fregene et al., 2000). Other duplicates, which are also morphologically very similar, were identified. These landraces were collected from South West and Central Nigeria and have different local names. The resistance in the landrace TME9, which also occurs in a separate clade on the phylogenetic tree, is likely to be CMD2. This landraces is found in the pedigree of IITA-TMS-961089A that was used in a parallel mapping study: resistance-linked markers also came from scaffold 05214 (Rabbi et al., in press). It is postulated that all of these landraces, which come from West Africa, trace back to a single CMD resistant
4.4. Screening of cassava mosaic geminiviruses in the mapping population

The CMG screening results showed that ACMV was the predominant species in the field, whereas EACMV occurred at lower frequency and only with ACMV as a dual infection. The preponderance of ACMV reflects the known proportions of the two viruses in West Africa and contrasts with the predominance of EACMV-UG in the pandemic regions of Eastern Africa (Legg and Fauquet, 2004). The severe form of EACMV-UG has displaced other CMG species and strains during the recent pandemic that swept through the region (Legg et al., 2006). An important question is whether the CMD2-type of resistance available in the West-African cassava germplasm can protect against these more virulent strains of CMG that does not contribute to the disease pressure in West Africa. Germplasm screening in Uganda has demonstrated that the qualitative resistance locus in the TME landraces is highly effective against this virus (Kawuki et al., 2011). Similar results have been obtained through biologic inoculation methods using-pseudo recombinants of ACMV and EACMV (Sserubambwe et al., 2008). Some of the screened clones include TME5 and TME14, which cluster together with TME3, the original source of the CMD2 locus (Fig. 5). The evidence suggests that CMD2 locus is a monogenic resistance with broad specificity against cassava-infecting geminiviruses.

4.5. Prospects for long-term efficacy of CMD2 resistance gene

Despite the apparent broad specificity of the CMD2 gene, whether the locus can confer durable resistance under classical definition (Johnson, 1984) depends on whether it will remain effective over a prolonged period of widespread use under conditions conducive to the disease. It is well known that monogenic resistance can be ephemeral and subjected to the well known boom-and-bust cycles (McDonald and Linde, 2002) as a result of the rapid genetic evolution in the pathogen though there is evidence of durable resistance from single gene actions (Johnson, 1984; Stuthman et al., 2007). Since its discovery in the 1980s and subsequent extensive use in breeding programmes in sub-Saharan Africa, there has been no report of breakdown of the CMD2-type of resistance in the field, indicating that the locus could be durable against CMGs. Still, the long-term effectiveness of the resistance locus needs to be augmented by combining it with the quantitative resistance derived from M. glaziovii. Indeed, crosses combining these two types of resistance have given rise to progeny which are near immune to all known CMG species, including the virulent EACMV-Uganda (Legg and Fauquet, 2004; Monde et al., 2012). Additionally, more efforts are needed for a comprehensive screening of the resistant landraces to identify any additional sources that may exist. Though increased resolution was achieved in the mapping analysis, a larger mapping population is required for fine-mapping and cloning of the concerned locus. This will provide insights on the mechanism of the resistance that so far seems to be highly effective against various CMG species.

Acknowledgments

This research was supported by the CGIAR-Research Programme on Roots, Tubers and Bananas and the “Next Generation Cassava Breeding Project”, through funding from the Bill & Melinda Gates Foundation and the Department for International Development of the United Kingdom. We acknowledge the help of Charlotte Acharya for preparation of GBS libraries and sequencing, and Oluwafemi Alaba for sample preparation and DNA extraction.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.virusres.2013.12.028.

References


