Theor Appl Genet (2015) 128:1839–1854
DOI 10.1007/s00122-015-2551-8

ORIGINAL PAPER

Fine mapping of *Msv1*, a major QTL for resistance to Maize Streak Virus leads to development of production markers for breeding pipelines

Sudha K. Nair[1] · Raman Babu[1] · Cosmos Magorokosho[2] · George Mahuku[3] · Kassa Semagn[3] · Yoseph Beyene[3] · Biswanath Das[3] · Dan Makumbi[3] · P. Lava Kumar[4] · Michael Olsen[3] · Prasanna M. Boddupalli[3]

Received: 6 December 2014 / Accepted: 30 May 2015 / Published online: 17 June 2015

Abstract

Key message ***Msv1*, the major QTL for MSV resistance was delimited to an interval of 0.87 cM on chromosome 1 at 87 Mb and production markers with high prediction accuracy were developed.**

Abstract Maize streak virus (MSV) disease is a devastating disease in the Sub-Saharan Africa (SSA), which causes significant yield loss in maize. Resistance to MSV has previously been mapped to a major QTL (*Msv1*) on chromosome 1 that is germplasm and environment independent and to several minor loci elsewhere in the genome. In this study, *Msv1* was fine-mapped through QTL isogenic recombinant strategy using a large F_2 population of CML206 × CML312 to an interval of 0.87 cM on chromosome 1. Genome-wide association study was conducted in the DTMA (Drought Tolerant Maize for Africa)-Association mapping panel with 278 tropical/sub-tropical breeding lines from CIMMYT using the high-density genotyping-by-sequencing (GBS) markers. This study identified 19 SNPs in the region between 82 and 93 Mb on chromosome 1(B73 RefGen_V2) at a $P < 1.00E-04$, which coincided with the fine-mapped region of *Msv1*. Haplotype trend regression identified a haplotype block significantly associated with response to MSV. Three SNPs in this haplotype block at 87 Mb on chromosome 1 had an accuracy of 0.94 in predicting the disease reaction in a collection of breeding lines with known responses to MSV infection. In two biparental populations, selection for resistant *Msv1* haplotype demonstrated a reduction of 1.03–1.39 units on a rating scale of 1–5, compared to the susceptible haplotype. High-throughput KASP assays have been developed for these three SNPs to enable routine marker screening in the breeding pipeline for MSV resistance.

Communicated by M. Frisch.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-015-2551-8) contains supplementary material, which is available to authorized users.

✉ Sudha K. Nair
Sudha.nair@cgiar.org

[1] International Maize and Wheat Improvement Center (CIMMYT), ICRISAT Campus, Patancheru, Greater Hyderabad 502324, India

[2] International Maize and Wheat Improvement Center (CIMMYT), Mount Pleasant, MP163 Harare, Zimbabwe

[3] International Maize and Wheat Improvement Center (CIMMYT), ICRAF Campus, UN Avenue, Gigiri, Nairobi, Kenya

[4] International Institute for Tropical Agriculture (IITA), Ibadan PMB5320, Nigeria

Introduction

Maize streak virus (MSV) disease is one of the major biotic constraints in the maize production systems of sub-Saharan Africa (SSA). MSV disease outbreaks often coincide with drought periods or irregular early rains exacerbating crop failures and often resulting in complete crop losses (Guthrie 1978). Despite the restricted geographical spread of MSV in sub-Saharan Africa (SSA), it was considered to be the third most important disease of maize in the world after gray leaf spot (GLS) and northern corn leaf blight (NCLB) (Pratt and Gordon 2006). Martin and Shepherd (2009) reported losses to the tune of US$120–480 million per year due to MSV in Africa in terms of lost income and higher maize prices, and indicated that at least half of such loss could be potentially recovered with the effective control of MSV. The International Maize and Wheat

Improvement Centre (CIMMYT), in collaboration with many national and international agricultural research institutions in SSA, has been successful in developing improved maize germplasm with resistance to MSV disease; several drought tolerant and MSV-resistant maize cultivars derived out of this germplasm have been released over the last 10–15 years and are reaching smallholders through the seed companies. Moreover, the varietal registration process in several countries in SSA requires that the new maize varieties nominated by the institutions possess reasonable levels of resistance to the MSV disease.

MSV belongs to the genus Mastrevirus and family Geminiviridae, and is obligately transmitted by as many as six leaf hopper species in the genus *Cicadulina*, and mainly by *C. mbila* and *C. storeyi.* MSV epidemiology is related to environmental influences on the vector species, leading to erratic epidemics in every 3–10 years (Martin and Shepherd 2009). The virus has a single-stranded circular DNA with approximately 2700 base pairs and is reported to have 11 strains of which, MSV-A is the most common and the cause for severe forms of the disease (Martin et al. 2001).

One of the most sustainable ways of managing the MSV disease is through host plant resistance. Maize breeders in SSA have been continually developing resistant varieties that are routinely deployed in the region. The first reports of resistance to MSV were in Peruvian Yellow and Arkkels Hickory cultivars, and an incompletely dominant gene was reported in the cross between these two cultivars (Storey and Howland 1967). The International Institute of Tropical Agriculture (IITA) researchers detected MSV resistance in tropical zea yellow (TZY) population, which has been used as a source of resistance (Soto et al. 1982). Genetic analysis of resistance to MSV in a highly resistant inbred line IB32 (derived from TZY) showed quantitative inheritance with the involvement of a few major genes (Kim et al. 1989). Rodier et al. (1995) reported both major and minor genes/loci governing MSV resistance, with complete to partial dominance of resistance. IITA and CIMMYT have developed several lines with MSV resistance (Kim et al. 1987; Wambugu and Wafula 2000), which are widely used by various institutions in SSA.

Quantitative trait loci (QTL) mapping for resistance to MSV was undertaken by several researchers (Kyetere et al. 1999; Welz et al. 1998; Pernet et al. 1999a, b; Lagat et al. 2008; Babu et al. Manuscript in prep). Based on an analysis of responses of maize genotypes to MSV in Uganda and Zimbabwe, Kyetere et al. (1999) identified a major gene on chromosome 1 (*Msv1*) in the tolerant line, Tzi4 (from IITA). Welz et al. (1998) revealed a large effect QTL on chromosome 1 in another MSV-resistant line (CML202 from CIMMYT) and suggested possible involvement of three additional minor QTL influencing MSV resistance. Pernet et al. (1999a, b) studied MSV resistance in two different resistant lines from Reunion Island, D211 and CIRAD390, and mapped a major QTL on chromosome 1. Babu et al. (Manuscript in prep), through an analysis of three MSV-resistant lines from CIMMYT (CML202, CML206 and MSRxPool9c1F1-176-4-1-4-Sn) at Harare (Zimbabwe), identified a major effect QTL on chromosome 1, along with moderate to minor effect QTL for MSV resistance on other chromosomes. Thus, several studies so far demonstrated the presence of a major effect QTL on chromosome 1 (hereafter referred to as *Msv1*), which seems very consistent across different genetic backgrounds and environments, besides several moderate to minor effect QTL that appear to be mostly germplasm-specific and/or environment-dependent. (Abalo et al. 2009) carried out marker-assisted selection (MAS) for MSV employing a SSR marker in the chromosomal bin 1.04 where the *Msv1* was reported consistently, to study the comparative economic advantage of MAS over conventional phenotype-based selection; the study concluded that MAS could be more economical compared to the latter.

CIMMYT, in partnership with several public and private sector institutions, is working to develop and deploy improved maize germplasm which are tolerant to abiotic stresses like drought, heat and poor soil fertility. The success of such germplasm in SSA largely depends on certain adaptive traits such as resistance to MSV and Maize lethal necrosis (MLN). Identifying and deploying genomic regions conferring resistance to MSV as well as for other important biotic stresses/adaptive traits will greatly accelerate the efforts towards rapid development and deployment of climate-resilient maize germplasm in SSA (Cairns et al. 2013). With the availability of tropicalized haploid inducers and doubled haploid (DH) facility in Africa (Prasanna et al. 2012, 2014), coupled with breeder-ready markers for MSV and other important adaptive traits, it is possible to significantly enhance genetic gains in maize and breeding efficiency in SSA.

For marker-assisted breeding to be successful, the confidence interval of the identified QTL needs to be as small as possible, with the best marker being the functional polymorphism responsible for the trait differences. In that case, the marker–trait associations are expected to be retained across diverse genetic backgrounds and generations. Coarse QTL mapping studies generally result in delimitation of large genomic intervals, which is not efficient for MAS due to rapid linkage disequilibrium (LD) decay in diverse tropical germplasm. Fine mapping involves mapping markers at sub-centimorgan (cM) distances from a gene/QTL of interest and is achieved through marker-based screening of a large number of progenies from either bi/multi parent population(s), followed by phenotyping of the marker-identified critical recombinants. With the onset of high-density genotyping platforms, such as genotyping-by-sequencing

(GBS) (Elshire et al. 2011), genome-wide association studies (GWAS) in collection of unrelated individuals can also potentially result in finer delimitation of large effect loci influencing a trait of interest. Historical recombinations that have occurred during the evolution of a crop species aid in achieving higher mapping resolution in GWAS (Myles et al. 2009). Mapping strategies that combine the advantages of both GWAS as well as linkage-based mapping help in comprehensive identification of causal loci along with simultaneous validation and finer delimitation of large effect loci of interest (Nordborg and Weigel 2008; Yu et al. 2008). The objectives of the present study were to fine-map the large effect QTL (*Msv1*) on chromosome 1 influencing MSV resistance, using a large number of progenies derived from a biparental population and an uncontrolled association population of tropical/subtropical breeding lines, and to develop a set of high-throughput production markers that could routinely be employed in the breeding programs for efficient incorporation of MSV resistance.

Materials and methods

QTL mapping

DNA extraction and genotyping

Genomic DNA was extracted from 3–4-week-old plants collected in a bulk of 15 plants per $F_{2:3}$ family according to CIMMYT's laboratory protocols (CIMMYT 2001). SNP markers were selected from the Illumina 1536 random SNP chip (Yan et al. 2009), HapMap_V2 (http://www.Panzea.org) and Illumina SNP50 (http://www.illumina.com). A total of 156 SNP markers spread across the genome were used for QTL analysis. All the SNP markers used in this study were genotyped as KASP assays from LGC genomics. (http://www.lgcgenomics.com).

QTL analysis

A F2:3 mapping population from a cross between CML206 (resistant to MSV) and CML312 (susceptible to MSV) (Babu et al. manuscript in prep) was used for mapping QTL for resistance to MSV. A linkage map was constructed with 156 polymorphic SNP markers using QTL IciMapping v3.3 software (http://www.isbreeding.net) where a LOD score of 3.0 and a maximum recombination frequency of 0.40 were used to declare linkage between two markers. QTL were identified based on the adjusted means of 253 $F_{2:3}$ families using an inclusive composite interval mapping (ICIM) implemented through QTL IciMapping v3.3. Walking step in QTL scanning was 1 cM, and a LOD threshold of 3.0 was chosen for declaring putative QTL. Additive (a) and dominance (d) effects for each QTL were estimated using QTL IciMapping v3.3. The type of gene action in the F_2 generation was determined on the basis of the dominance ratio (DR = |2*d*/*a*|) as described by Stuber et al. (1987): additive for DR < 0.2, partial dominant for $0.2 \leq$ DR < 0.8, dominant for $0.8 \leq$ DR < 1.2 and overdominant for DR $\geq$ 1.2, where *d* had to be multiplied by 2 as it was estimated from F_3 families.

Fine mapping of *Msv1*

Plant materials and selection of QTL isogenic recombinants (QIRs)

A QIR strategy (Peleman et al. 2005) was adapted for fine mapping *Msv1*. A large F_2 population consisting of 4725 F_2 seeds was developed from CML206 X CML312 and seed DNA extraction was carried out, based on the procedure suggested by Gao et al. (2008). Genotyping was carried out using the two flanking markers identified in QTL analysis, PZA00944.1 and CSU11338.4, along with two additional markers, PZE0165109124 and PZE01141571676, farther to the flanking markers identified for the *Msv1* interval. The context sequences of the markers are given in Table 1. Recombinants which were heterozygous at one of the flanking markers and homozygous at the other flanking marker were selected, and only the selected seeds were planted in the field. These plants were selfed and 12 seeds from each selfed ear were again genotyped using PZE0165109124 and PZE01141571676 to identify homozygous recombinants. The homozygous recombinant genotypes were further analysed using flanking markers of QTL identified on chromosomes 3 and 10 (QTL3 and QTL10, respectively). Markers used to characterise QTL3 were PZA01396.1 and PHM17210.5, while markers used to characterise QTL10 were PZA01919.2 and PHM13687.14. Two categories of QTL isogenic recombinants (QIRs) were selected from among the homozygous recombinant individuals, with the first category having resistant alleles both at QTL3 and QTL10 and the second category having susceptible alleles both at QTL3 and QTL10 (Fig. 1). Along with QIRs, four sets of control plants with the following genotypic constitution at the three QTL were also selected, (1) susceptible alleles at *Msv1*, QTL3 and QTL10, (2) resistant alleles at *Msv1*, QTL3 and QTL10, (3) resistant allele at *Msv1* and susceptible alleles at QTL3 and QTL10 and (4) susceptible allele at *Msv1* and resistant alleles at QTL3 and QTL10. The QIRs in each category were classified as resistant or susceptible based on the phenotypic range of the corresponding control sets. If the phenotypic value of a particular QIR falls outside the range of the control sets, they were not further included in the analysis.

Table 1 SNP markers used in fine mapping of *Msv1* with their context sequences

SNP markers	Physical position on chr.1 (B73 RefGen_V2)	Context sequences
PZE0165109124	65087499	CTAATTGCGTTGTTGACTGTAGTTAGTATTYTTATTGCCTATACCGCAGC[T/G] AACAGTAGGCATATTCTGTCTCTCAACTTTGATAGCAYTCAATATCAGTT
PZE-101084671	73349070	CACCCAACAATACTATTTAAGGTCTTTGCCTATCCTTTTTCGTTGTCCAA[A/G] CTACCAGCAGCTATGCGTATCGCTTTTGGTGATGCCTGTATGCATAGTTG
SYN35629	81106413	GTTGGCCGTGGAGTTCAAGTAGGCCAGGATCTTGGACCCGGACTCGTAGC TGACGTGCGA[A/C]GCCGGGAGGACGTGCGCGTCGGCGAAGGTGGTGTAC CCCTCCGCCTCCTTGTTCATGACG
PZE-101090728	82577010	GCTGAGACGATGTTCTTGAACCAAGCTCCCTGGAAACTAGGGCTGCCTCT[A/G] TTTTGATTGTTACCCGGAGACTCAGGTGAGGCTTGATTTTTGGAAGTCAG
SYN25468	83783583	GTTGGCCGTGGAGTTCAAGTAGGCCAGGATCTTGGACCCGGACTCGTAGC TGACGTGCGA[A/C]GCCGGGAGGACGTGCGCGTCGGCGAAGGTGGTGTAC CCCTCCGCCTCCTTGTTCATGACG
PZE-101093951	87301345	TAACTCTCTGCTGTTGCTTGTCTTCAGGTTGTCATGAGAGATCCTCACAT[A/G] GCAGCAGATGGCTTCACCTACGAAGCTGACGCTCTTAGATACTGGCTCGA
PZE0186065237	87301459	ACATCTCCAGTAACAAACAGAAGTCTTTCGAATCGTGATACCATCCCCAA[T/C] CACGCACTGCGKTCGGCCATCCAAGAATACCTCCGGCAGAACGAGCTGCA
PZE0186365075	87601397	AGAAGAAAATGGCCTGCCATATATATATCCCGGTTAATCGCTARTGCATT[A/C] TCAGGAATCATTCTCATAGGTCATAAGACGAGCAAGGGATACTCTTCTAC
PZE-101098295	89384498	CACACATGGTTGGGATTAGACCCTAGCATACCCCTTCGACTATTTAATCT[A/G] AGTCGTGGATAGTTGAACGGGCGGTTCCGGTTAGAGATAAAGTAACGCTT
PZE-101098418	89771326	TGCCGAGTGCTATGGCCATGACACTCGACAAAGTACCTTACTGTGAACAC[A/G] CCAGGCGTAGAATAGCCCACACGCCGGTAAGTCATGCATGGAGTACTGGC
PZA00944.1	91429024	CAAATGAGGTGCCACTTCGGGTGGAAAAATATGCTTGTATAGTGGAT CGACACGGTTGTC[A/G]TCCACGGTGATAGTGCTGTCATCTCCAAGGAAT GGAAAGTCATACAGGTAGCTAGGCAAC
PZE-101094951	92871751	AAAATAAAAGAATGGGCACGACATAAATTACCTAGGGTCATGGCGGCGCA[A/C] GCCCATCCGCCGGTGGTGGCACGGGAAGCAGGACGACCCTTGTAGTCCCA
PZE-101102819	102090147	AAATAGGGCGGTTGGGTAGTACCAAATCCTATCTGATCCTTGTAGGATGA[A/G] AAGGGCTCAGGAGCGTCATCCTTGCTGGGCGGAGCGACGTCCCTGACGAG
PZE-101107639	113004922	GGGGGCCGAGGAATCCTTCGCATTAGCTGAGCCGCCATCGCCAGGACTCA[C/G] CCTGCATTCACCCTCGATGAATGAGACACGTCCTCCATGACCCTCTACTG
CSU1138.4	119018556	AAAARGCRGGACTGCAAACCACGYTRGCTAGACCACACATGCAACTCTNNN NTAGCTAGC[A/G]ACGTCGATCACTGACCACCATTCAACTCCATCGATCAA TGCTTGGCCGGAGGGTGGTAGA
PZE01141571676	142730137	TACCTGGTGAGCGCGTCCAGGACGACGACGGCCATGCCGCGGTTGTCCCC[G/A] CGGGAAGTCTTGGGCTGGTTGTCGCCCTTCATGGAGACGTCGTCGTAGAA

Phenotyping of QIRs and control sets

A total of 126 entries consisting of 122 selected $F_{2:3}$ lines comprising QIRs and control plants, their parental lines CML206 (MSV-resistant) and CML312 (MSV-susceptible), along with two standard checks (CML312 and CML539) were planted in an alpha lattice design with two replications at CIMMYT experimental station at Harare in Zimbabwe in 2012 and 2013. Plots consisted of single rows, 2 m long, with distances of 20 cm between plants and 75 cm between rows. Artificial infestation with viruliferous leaf hoppers (*C. mbila*) was conducted once in each trial, using standard procedures (Tang and Bjarnason 1993). MSV severity was rated twice on 10 individual plants per plot, using a 1–5 scale with half points, as follows: 1 = no symptoms on leaves; 1.5 = very few streaks on leaves; 2 = light streaking on old leaves, gradually decreasing on young leaves; 2.5 = light streaking on old and young leaves; 3 = moderate streaks on old and young leaves; 3.5 = moderate streaks on old and young leaves and slight stunting; 4 = severe streaking on 60 % of leaf area and plants stunted; 4.5 = severe streaking on 75 % of leaf area and plants severely stunted; and 5 = severe streaking on 75 % or more of the leaf area, plants severely stunted or dead.

Variance components for individual and combined trials were estimated by linear mixed model analysis using the PROC MIXED of SAS (REML option) as implemented

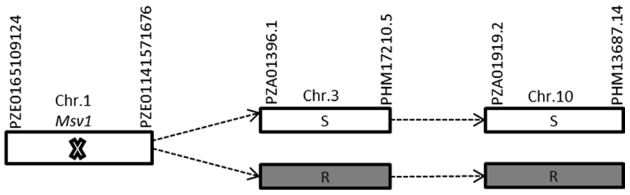

Fig. 1 Strategy for selection of QIR lines from a large F_2 population. QIR lines are selected from large F_2/F_3 population based on recombination in the *Msv1* genetic interval, and the other QTL regions are fixed as either genotypically resistant or susceptible types. The flanking markers used in this analysis for the three QTL regions are depicted here

in the CIMMYT field book. For individual analysis, broad sense heritability (repeatability) was calculated as $h^2 = \sigma_g^2/[\sigma_g^2 + \sigma_g^2/r]$ and for combined analysis, as $h^2 = \sigma_g^2/[\sigma_g^2 + \sigma^2 gy/y + \sigma^2 e/yr]$, where σ_g^2 is the genotypic variance, σ_g^2 is the error variance, y is the number of years and r is the number of replications in each trials. The heritability estimates in this study could have been over-estimated in the single and combined analysis due to lack of multiple seasons and locations, respectively, in the model.

Marker analysis

To resolve the exchange boundaries of recombination events within the Msv1 interval, the selected QIRs were genotyped with fourteen additional SNPs that mapped in the flanking marker interval of PZE0165109124 and PZE01141571676. The SNPs were selected from HapMap_V2 and MaizeSNP50 Genotyping BeadChip from Illumina based on the physical location corresponding to B73 RefGen_V2. KASP assays were developed at LGC genomics for the selected SNPs and were assayed on the QIRs after validation.

GWAS and LD-based haplotype analysis

An association mapping panel developed for the Drought Tolerant Maize for Africa project (DTMA-AM) comprising 278 lines, mostly from CIMMYT's tropical and sub-tropical breeding programs (Cairns et al. 2013), was evaluated for responses to MSV under artificial infection at CIMMYT-Harare in 2010. The DTMA-AM lines were genotyped using GBS at the Institute of Genomic Diversity, Cornell University as well as with the SNPs surrounding the *Msv1* interval used in fine mapping. The original dataset comprised ~1 million SNPs from which markers with more than 0.3 Call Rate (CR) and 0.02 Minor Allele Frequency (MAF) were included for GWAS. For calculating PCA and kinship matrix, a subset of high-quality SNPs (CR > 0.9 and MAF > 0.1) was used.

A mixed linear model (MLM) analysis was employed for GWAS. MLM is a regression analysis of the trait on the genotypic data while correcting for cryptic relatedness and pedigree structure (Yu et al. 2008). Individual SNP-based association tests were conducted using the MLM procedure as implemented in SNP & Variation Suite (SVS) V_7.7.8 (SVS, Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). The association mapping model used was $Y = \text{SNP}^*\beta + \text{PC}^*\alpha + \text{K} + \varepsilon$, where, Y = response of the dependent variable (MSV disease score), SNP = SNP marker (fixed effects), PC = principal component coordinate from the PCA (fixed), K = kinship matrix (random), α and β are SNP and PC fixed effect model coefficients, respectively, and ε is the error. Principal component analysis (PCA) was performed in SVS and the first ten principal components were used as covariates in the linear models. A kinship matrix was also computed from identity-by-state distances matrix using SVS as

$$\text{IBS distance} = (\text{no. of markers IBS2}) + 0.5 \times (\text{no. of markers IBS1})/\text{no. non-missing markers},$$

where IBS1 and IBS2 are the states in which the two inbred lines share one or two alleles, respectively, at a marker (Bishop and Williamson 1990). The kinship analysis was also carried out using SVS.

LD analysis and haplotype trend regression

The extent of genome-wide LD was evaluated based on adjacent pairwise r^2 values (the squared correlation coefficients among alleles at two adjacent SNP markers) between 50,000 randomly distributed SNPs assayed by GBS across all chromosomes and physical distances among these SNPs. Non-linear models with r^2 as response (y) and pairwise distance (x) as predictors were fitted to the genome-wide LD data using the 'nlin' function in R package (R core team 2014). Average pairwise distances in which LD decayed at $r^2 = 0.2$ and $r^2 = 0.1$ were then calculated based on the model. The expected value of r^2 was

$$E\left(r^2\right) = \left[\frac{10 + C}{(2 + C)(11 + C)}\right]\left[1 + \frac{(3 + C)(12 + 12C + C^2)}{n(2 + C)(11 + C)}\right],$$

where r^2 = squared correlation coefficient, n = sample size and C is a model coefficient for the distance variable (Hill and Weir 1988).

For calculating the LD in a specific region on chromosome 1, the GBS SNPs in the physical interval were extracted. LD analysis was performed on all adjacent pairs within the selected region. Both r^2 and D′ values were calculated using the Expectation Maximisation (EM) algorithm (Excoffier and Slatkin 1995) as implemented in SVS. Average adjacent pairwise LD (r^2) was calculated between

Table 2 Genetics of QTL mapped for MSV resistance in the $F_{2:3}$ population from CML206 X CML312

Chromosome	Position	Leftmarker	Rightmarker	LOD	Est add	Est dom	PVE %	Action
1	131.5	PZA00944.1	csu1138.4	11.76	−0.983	−0.2358	67.36	PD
3	225.7	PZA01396.1	PHM17210.5	3.46	−0.355	−0.0429	8.6	PD
10	21.2	PZA01919.2	PHM13687.14	4.87	−0.139	0.5671	10.4	OD

PD partial dominance, *OD* over dominance

all the SNPs in the region. Haplotype frequencies were estimated from among the 20 SNPs within 82 Mb to 93 Mb associated with the trait at a *P* value < 1E-04 using 50 EM iterations, an EM convergence tolerance of 0.0001 and frequency threshold of 0.01. Haplotype blocks were detected based on the block defining algorithm to minimise historical recombinations (Gabriel et al. 2002). Haplotype trend regression was carried out based on a stepwise regression of MSV phenotype with the pre-estimated haplotypes with backward elimination and a P-value cut-off of 0.001.

Validation in biparental populations

Two biparental populations were developed and phenotyped for MSV disease reaction at Harare (Babu et al. Manuscript in prep). 258 F_3 families from [MSRxPool9]c1F1-176-4-1-4-Sn (MSV-resistant) X CML312 (MSV-susceptible) and 252 F_3 families from CML202 (MSV-resistant) X CML312 (MSV-susceptible) were evaluated. Each of the F_2 plants forming these families were genotyped with KASP assays developed for three SNPs that form the most significant haplotype associated with MSV disease reaction. The F_3 families were classified as belonging to the resistant parent type or susceptible parent type haplotypes. The t-test and ANOVA were carried out using the R package (R Core Team 2014).

Results

QTL mapping for MSV resistance

An inclusive composite interval mapping carried out on 253 $F_{2:3}$ families identified three QTL conferring resistance to MSV. A major QTL which explained 67.4 % of the phenotypic variance was detected on chromosome 1 between markers PZA00944.1 and CSU1138.4, within a genetic distance of 5.9 cM (Table 2). This major QTL falls in the same bin on chromosome 1 (1.04/1.05) where *Msv1* was reported in various earlier studies. For this reason, it is assumed to be allelic or identical to *Msv1*, and will henceforth be referred to by this name. Apart from this QTL, two other QTL were identified on chromosomes 3 and 10, which explained 10.4 and 8.6 % of the phenotypic variance. The estimated R^2 values could have been over-estimated as no cross-validation was carried out as part of the QTL mapping in the ICIM software. All three QTL detected were contributed by the resistant parent, CML206. The major QTL on chromosome 1 and the QTL on chromosome 3 were found to be partially dominant, whereas the one on chromosome 10 was found to be over dominant according to the classification by Stuber et al. (1987).

Identification of QTL isogenic recombinant (QIR) sets

Msv1 was fine-mapped using the QTL isogenic recombinant strategy as described in the "Materials and methods". Initial marker analysis was carried out on the large F_2 population of CML206 × CML312 with six markers flanking the three QTL identified. Additional SNPs farther to the flanking markers, PZE0165109124 and PZE01141571676 were genotyped in the large F_2 population, to consider informative recombination events which could have occurred immediately outside the delimited intervals of the QTL. 395 recombinants were selected between markers PZE0165109124 and PZE01141571676, which were homozygous at one marker and heterozygous at the other. We raised F_3 progenies from these heterozygous recombinants to select for homozygous recombinant individual progeny that are isogenic for the other two QTL identified in this population. This was done as a refinement to the original QIR strategy in order to further multiply selected QIRs and to facilitate precise phenotyping across multiple seasons/locations. From among these, 36 QIRs were selected having either homozygous susceptible genotypes for QTL3 and QTL10 or homozygous resistant genotypes for QTL3 and QTL10. Additionally, control sets were derived by selecting homozygous susceptible genotypes for the three QTL; homozygous resistant genotypes for the three QTL; homozygous susceptible genotype for *Msv1*, but homozygous resistant types for QTL3 and QTL10 and homozygous resistant genotype for *Msv1*, but homozygous susceptible genotypes for QTL3 and QTL10 (Fig. 1). The entire set of QIRs and control sets were phenotyped in 2012 and 2013 under artificial MSV infection at CIMMYT-Harare.

Phenotyping of QIRs and control sets

The set of 122 lines selected from the F_3 population, as described above, were evaluated under artificial MSV

Table 3 (a) Estimates of mean, variance components and heritability of QIR and control set trials for MSV reaction, (b) summary statistics on MSV reaction of different QIR and control sets formed for QIR analysis

Year	Mean	Min	Max	σ_g^2	σ_g^2	σ_{gy}^2	*h*2	CV
3a								
2012	2.54	1.12	4.19	0.16	0.18		0.64	20.80
2013	2.67	1.27	4.27	0.45	0.18		0.83	11.30
Combined	2.62	1.20	4.25	0.37	0.29	0.05	0.78	

Type	*n*	Min	Max	Mean	SE	SD
3b						
S QTL1: S QTL3: S QTL10	20	3.00	4.27	3.38	0.11	0.35
R QTL1: S QTL3: S QTL10	22	2.09	2.70	2.27	0.04	0.12
R QTL1: R QTL3: R QTL10	20	1.20	2.00	1.62	0.08	0.27
S QTL1: R QTL3: R QTL10	24	2.35	3.28	2.71	0.09	0.32
Rec QTL1: S QTL3: S QTL10	19	2.10	3.95	2.97	0.17	0.66
Rec QTL1: R QTL3: R QTL10	17	1.33	3.26	2.31	0.16	0.59

3a: σ_g^2 genotypic variance, σ_g^2 error variance, σ_{gy}^2 interaction variance (genotype × year)

3b: The shaded text indicate the two QIR categories a and the italicised test indicates the control sets

inoculation at CIMMYT-Harare as previously described. The set comprised 86 families from four different control sets and 36 families representing QIRs derived based on recombination events from the large F_2 population as described above. The heritability estimates were 0.64 and 0.83 in 2012 and 2013, respectively, and 0.78 for the combined analysis (Table 3a). The summary statistics pertaining to each set of families are given in Table 3b. The parental lines CML206 and CML312 showed an adjusted mean disease score of 1.64 and 4.00, respectively, indicating significant phenotypic contrast.

Fine mapping of *Msv1* by QIR analysis

We selected two types of QIRs, one set having recombination within the *Msv1* interval and resistant alleles at the other two QTL on chr3 and chr10. The MSV phenotypes of these lines were compared against the control sets having resistant alleles at all the three loci and the control sets having the susceptible allele at *Msv1* and resistant alleles at chr3 and chr10. The QIR set of this category had average phenotypes across years ranging from 1.33 to 3.26 (Table 3b). The control sets with resistant alleles at all the three QTL had a phenotypic range of 1.20 to 2.00, and the other control set having susceptible allele at *Msv1* and resistant alleles at QTL3 and QTL10 had a phenotypic range from 2.35 to 3.28. Out of the 17 F_3 families in this QIR category, only 14 could unambiguously be assigned genotype category of *Msv1* based on the phenotypes of the control sets (Fig. 2). Secondly, we selected a QIR set with recombination at *Msv1* and having susceptible alleles at QTL3 and QTL10. These were compared against control sets having susceptible alleles at all the three loci and control sets having resistant allele at *Msv1* and susceptible alleles at QTL3 and QTL10. The QIR set of this category had phenotypic values ranging from 2.10 to 3.95. The control set having susceptible alleles at all three loci had a phenotypic range from 3.00 to 4.27, whereas the control set with resistant allele at *Msv1* and susceptible alleles at QTL3 and QTL10 had a phenotypic range from 2.09 to 2.70. Of the 19 F_3 families in this QIR set, four could not be assigned a definitive genotype category for *Msv1* as they fell outside the range of the control sets they are compared against. Based on the control set phenotypes, the genotypes of *Msv1* in the families in this QIR set are depicted in Fig. 2.

Based on the genotypic categories derived based on the comparison between control sets with varying composition of resistant and susceptible genotypes at *Msv1* and the other two QTL, and analysing them against the recombination exchange boundaries in QIRs as discussed above, the *Msv1* interval was delimited to a 7.62 Mb interval, flanked by PZE-101090728 (82.15 Mb) and PZA00944.1 (89.77 Mb). There were four QIRs in the category having susceptible alleles at QTL3 and QTL10 and five QIRs in the category having resistant alleles at QTL3 and QTL10 (Fig. 2), which determined the exchange boundary of *Msv1*. Six additional markers were selected from among the HapMapV_2 and MaizeSNP50 within this physical interval corresponding to B73 genome sequence (B73 RefGen_V2), but they were not able to further refine the interval as five of the six markers co-segregated with PZE-101090728 and one marker co-segregated with PZA00944.1. The genetic distance of the delimited *Msv1* interval was 0.87 cM in the large F_2 population based on two-point linkage analysis.

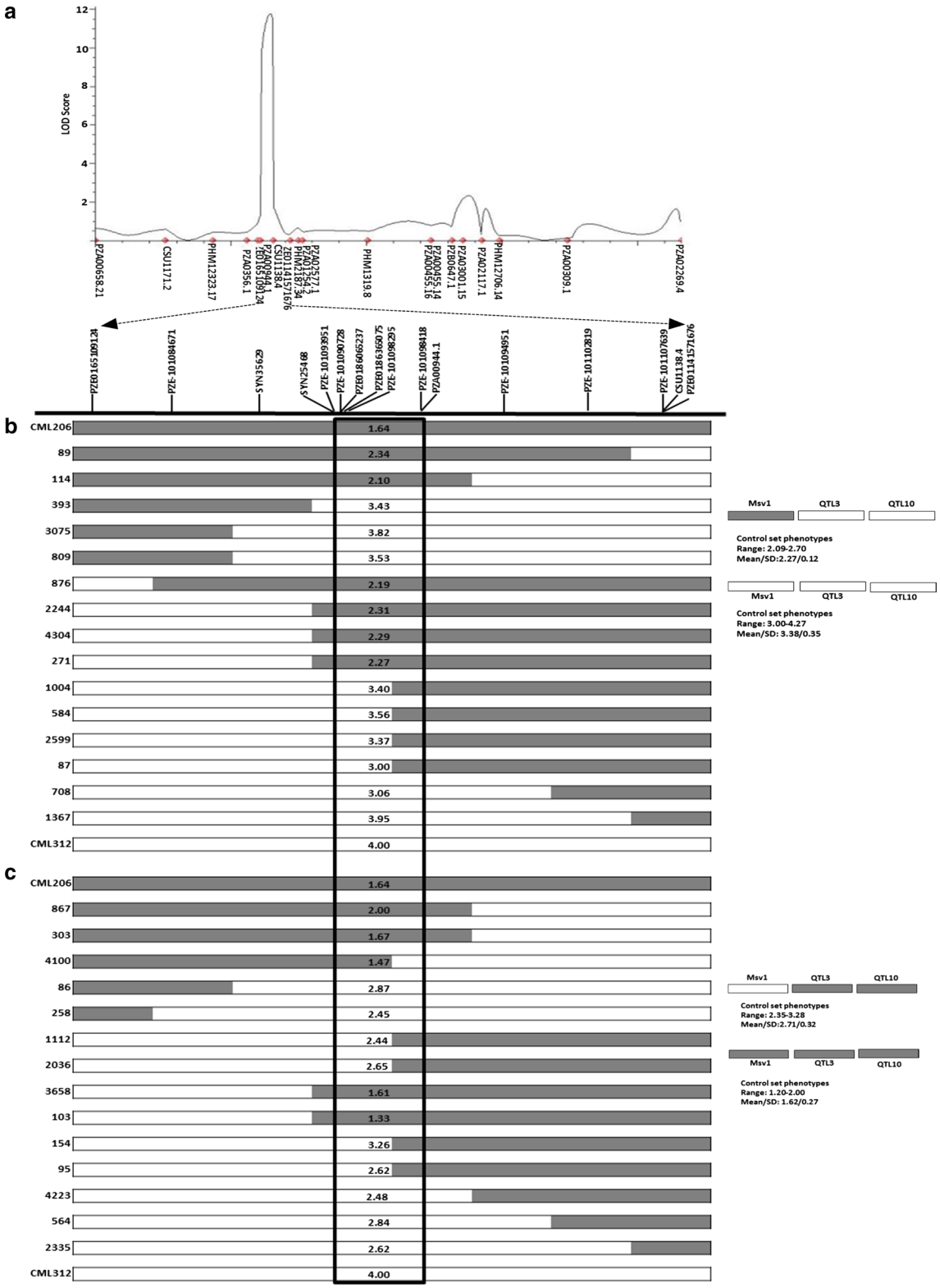

Fig. 2 Fine mapping of *Msv1* region employing the QIR strategy. The topmost panel shows the QTL interval on chromosome 1 identified through inclusive composite interval mapping. In Panel **b** and **c**, *filled bars* represent resistant alleles and *open bar*s represent susceptible alleles. The panel **b** shows QIR lines (entry numbers depicted in the *left side*), with susceptible alleles at QTL3 and QTL10. The panel **c** shows QIR lines (entry numbers depicted in the *left side*), with resistant alleles at QTL3 and QTL10. The adjusted means of all QIR lines are depicted at the centre of each bar in panel **b** and **c**. The *bold rectangle* in the panels **b** and **c** represents the delimited interval from the QIR analysis. The phenotypic controls for the QIR lines in panels **b** and **c** are represented in the *right side* of the respective panels

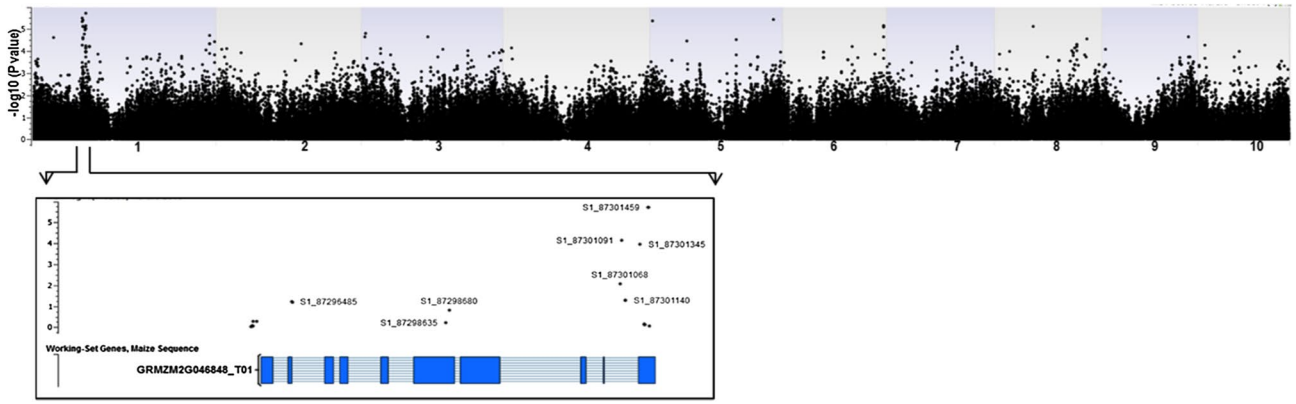

Fig. 3 The most significant SNP association to MSV disease reaction identified from GWAS (Q+K model) represented by a Manhattan plot, plotted with the individual SNPs on the X-axis and −log10 *P* value of each SNP derived from the association study in the Y-axis. Below, the genomic position of the identified SNP was retrieved from the working gene sets in B73 Ref gen_V2, as located in the last exon of GRMZM2G046848

GWAS for MSV resistance

The DTMA-AM panel comprises 278 diverse inbred lines developed at CIMMYT; these are adapted to the tropical and sub-tropical regions of SSA and Latin America (Cairns et al. 2013). The panel was phenotyped for responses against MSV under artificial infection in an alpha lattice design with two replications at CIMMYT-Harare. The heritability of the trial was 0.79. The disease rating ranged from 1.00 to 4.50, with a mean of 2.67 (Supplemental Fig. 1). Out of the total 955,960 SNPs obtained from GBS, various subsets were used to analyse the population structure, kinship, genome-wide LD and marker-trait associations.

For Mixed Linear Model (MLM) GWAS, principal components and kinship matrix was calculated using a subset of all the markers (146,444) which satisfied the stringent criteria of call rate (CR) $\geq$ 0.9 and minor allele frequency (MAF) $\geq$ 0.1. For the association analysis, a larger dataset included all SNPs that met the filtering criteria of CR $\geq$ 0.3 and MAF $\geq$ 0.02. A total of 337,518 SNPs fulfilled these criteria and on an average there was one marker at every 6.1 Kb. Based on individual SNP-based MLM association analysis correcting for population structure contributed by first 10 PCs and kinship, 58 SNPs were detected with highly significant associations at $P < 1.0E-04$ (Supplemental Table 1). Of these, 19 were in a region between 82 and 93 Mb on chromosome 1, coinciding with the fine-mapped physical interval of *Msv1* through QIR analysis, as described in the earlier section. The most significant SNP for MSV resistance (S1_87301459) was located at 87.30 Mb, incidentally within a gene, GRMZM2G046848 as per the B73 RefGen_V2 (Fig. 3). The minor allele of the SNP reduced the disease score, by 1.33 units and was present in about 32 % of the entire panel.

The LD within the 11 Mb physical interval was further studied by extracting 1405 SNPs from the region. The average distance between adjacent SNPs was 7.74 Kb, and the average adjacent pairwise LD (r^2) was 0.41, higher than the genome-wide adjacent pairwise average r^2 of 0.19, estimated using a random set of 50,000 SNPs across the genome. LD appeared to be high in this region relative to the whole genome. Similarly, LD in the 11 Mb region was found to decay ($r^2 = 0.1$) at an average distance of 1.61 kb compared to the whole genome average of 0.69 Kb, which indicated slower rate of LD decay in the *Msv1* interval. Among the 19 trait-associated SNPs within this interval, seven haplotype blocks were detected and their frequencies estimated using EM algorithm (Fig. 4). A haplotype trend regression was carried out by regressing MSV phenotypic scores of the DTMA panel lines onto the seven haplotype blocks. This analysis identified the 5th haplotype block (Haplo_5), composed of three SNPs viz., S1_87301345, S1_87301459 and PZE0186365075 as the most important determinant (r^2: 45.2 % with full model *P* value: 1.00E-28) of reaction to MSV infection (Table 4). Three haplotypes were observed among the DTMA-AM panel lines for this haplotype block -"GTA", "ACC" and "GTC", having EM probabilities of 0.613, 0.338 and 0.033, respectively.

KASP assays and their validation in breeding lines

The physical positions of two GBS SNPs that formed Haplo_5 coincided with SNP markers from MaizeSNP50 and HapMap_V2. (S1_87301345 with PZE-101093951, based on Maize SNP50; S1_ 87301459 with PZE0186065237 based on Hapmap_V2). The third SNP that formed the Haplo_5 block was PZE0186365075 from the Hapmap_V2. Based on the context sequences

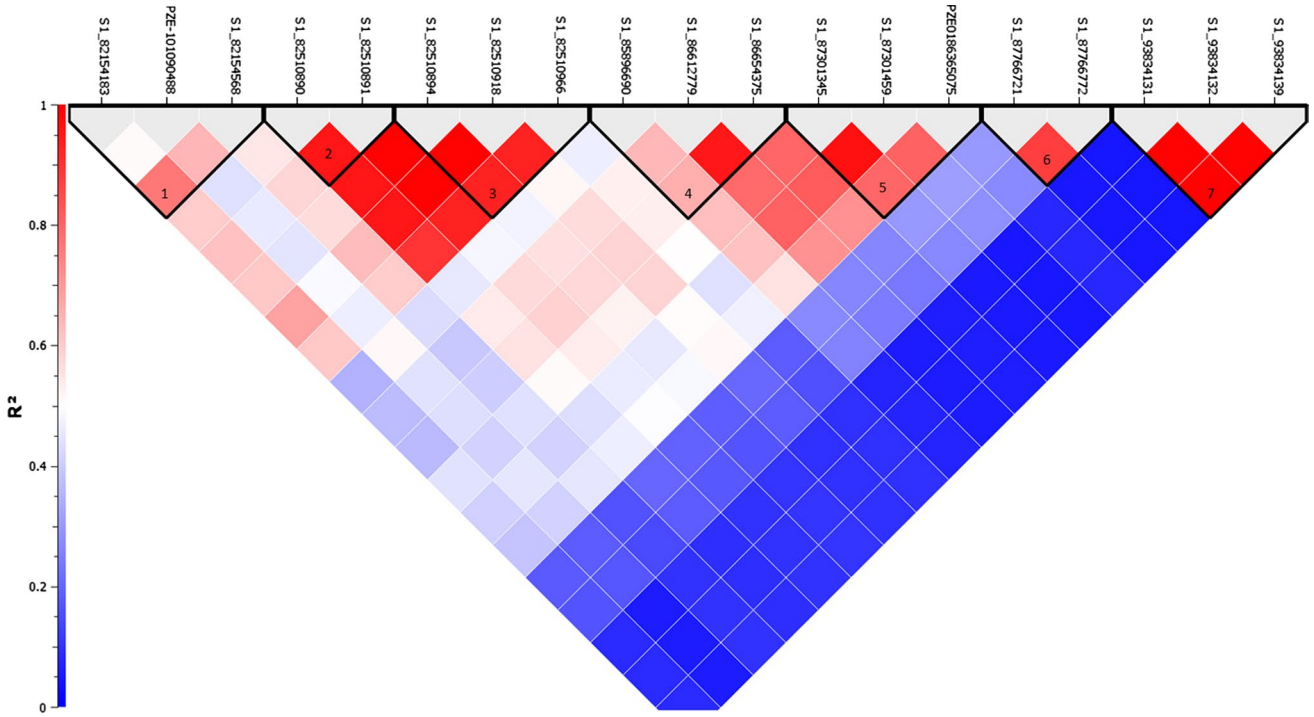

Fig. 4 Haplotype blocks detected in a region of 11 Mb between 82 and 93 Mb of chromosome 1 identified with significant SNP associations with MSV disease reaction. The seven haplotype blocks and the markers are depicted. The *vertical bar* on the left shows the legend for the strength of LD measured as r^2, with maximum LD depicted in red and least LD depicted in blue

Table 4 Haplotype trend regression of MSV responses of DTMA-AM panel lines with haplotypes identified in *Msv1* region identified from GWAS

Block#	# Haplotypes	Full model *P* value	−log10 full model P	r^2
1	5	9.393E−19	18.0272	0.384
2	2	1.577E−25	24.80216	0.354
3	3	4.139E−28	27.38315	0.407
4	4	7.58E−29	28.12029	0.405
5	3	1.003E−28	27.99876	0.452
6	3	1.723E−15	14.76376	0.216
7	2	6.466E−12	11.18936	0.163

of the three SNPs, we designed KASP assays and genotyped the QIRs. The three markers co-segregated with PZE-101090728, one of the flanking markers delimiting the *Msv1* interval. As the physical positions of these SNPs fall within the fine-mapped physical interval and are associated with MSV phenotype, we tested these markers for their efficiency as predictors for the trait. For this, a set of 89 CIMMYT Maize Lines (CMLs), with known responses to MSV infection (resistant/susceptible), were chosen along with 77 lines from the DTMA-AM panel which showed extremes of resistance/susceptibility (1.00–1.50 for resistance and 3.50–4.50 for susceptibility). This set of 166 elite breeding lines were genotyped with the KASP markers developed for the three SNPs identified. Out of the 166 lines, 92 were resistant and 74 were susceptible to MSV. Seven haplotypes were detected (Table 5), with the most prominent ones being Haplo_5.1 ("ACC") and Haplo_5.2 ("GTA") for the SNP markers PZE-101093951, PZE0186065237 and PZE0186365075, respectively. Haplo_5.1 was associated with resistance in 73 out of 76 occurrences and Haplo_5.2 was associated with susceptibility in 68 out of the 72 occurrences. The haplotypes Haplo_5.3 ("GTC") and Haplo_5.4 ("GCA") were associated with resistance in 8 out of 10 occurrences and 4 out of 4 occurrences, respectively (Supplemental Table 2). Haplo_5.1 could predict resistance 80 % of the times and Haplo_5.3 could predict resistance 9 % of the times. The remaining portion was explained by the rest of the haplotypes. Similarly, Haplo_5.2 could predict susceptibility 92 % of the times and the remaining 8 % was shared by Haplo_5.1, Haplo_5.3 and Haplo_5.7. Haplo_5.4, Haplo_5.5 and Haplo_5.6 occurred very rarely in the sample studied and were associated with the resistant phenotype. The most common haplotypes among a range of tropical/subtropical breeding lines, including the CMLs identified in this study could classify their MSV response phenotypes to a fairly high degree of precision.

Table 5 Haplotypes and haplotype frequencies in a set of elite CIMMYT breeding lines which shows distinct resistance or susceptible responses to MSV under artificial infection

Physical position (RefGen_V2) SNP	C1:87,301,345 PZE-101093951	C1: 87,301,459 PZE0186065237	C1: 87,601,397 PZE0186365075	Resistant	Susceptible	Total
Haplo5.1	A	C	C	73	3	76
Haplo5.2	G	T	A	4	68	72
Haplo5.3	G	T	C	8	2	10
Haplo5.4	G	C	A	4	0	4
Haplo5.5	A	C	A	1	0	1
Haplo5.6	A	T	A	1	0	1
Haplo5.7	G	C	C	1	1	2
						166

Validation of *Msv1* haplotype in independent biparental populations

The haplotype block associated with MSV resistance identified from our analyses was tested to determine what portion of the phenotypic variance this could explain in independently derived segregating progenies. Two previously phenotyped $F_{2:3}$ populations ([MSRxPool9] c1F1-176-4-1-4-Sn X CML312 and CML202 × CML312) were genotyped with the three selected KASP SNPs. The resistant parent in the first population had the haplotype "ACA" (Haplo_5.5), whereas the resistant parent in the second population had the haplotype "ACC" (Haplo_5.1). The susceptible parent had the haplotype "GTA" (Haplo_5.2). An ANOVA was performed after partitioning the $F_{2:3}$ genotypes into the two parental haplotype classes. In the population [MSRxPool9] c1F1-176-4-1-4-Sn X CML312, there were 52 families in the resistant parent haplotype group and 54 families in the susceptible parent haplotype group. The mean MSV disease scores of the resistant and susceptible type families were 2.05 and 3.44, respectively, with the *t* test *P* value < 2.2e-16 (Fig. 5). For the population CML 202 X CML312, the 51 resistant and 47 susceptible type F_3 families had the mean MSV disease scores of 2.26 and 3.29, respectively. The ANOVA in both populations showed that variation for this haplotype has a highly significant effect on MSV score, with the variance explained due to the haplotypes being 37 and 26 % in the two populations, respectively.

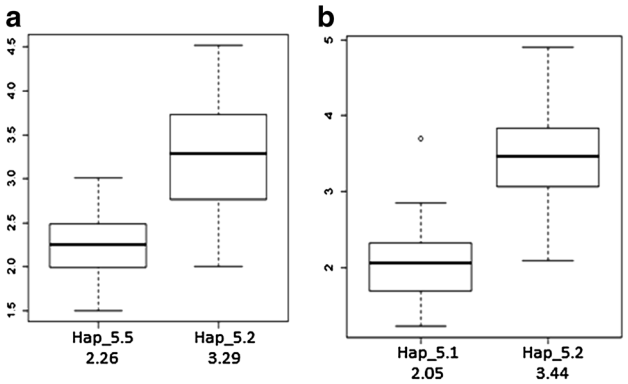

Fig. 5 Box plots depicting the MSV scores of $F_{2:3}$ families classified based on genotypically resistant haplotype type and genotypically susceptible haplotype type at *Msv1* in, **a** CML202 X CML312, and **b** [MSRxPool9] c1F1-176-4-1-4-Sn X CML312. The *boxes* represent the first and third quantiles and the median is represented by a short *black line* within the box. The lines extending from the boxes to the *horizontal bars* represent the distance to the maximum and minimum observations. Outliers are represented by isolated *circles*

Discussion

Developing improved maize cultivars with genetic resistance to MSV is an important component of sustainable crop management strategy in SSA. International institutions, such as CIMMYT and IITA, have partnered with several regional and national institutions to develop and deploy an array of maize hybrids and OPVs with enhanced levels of MSV resistance in SSA, primarily through conventional breeding. Despite the significant success, phenotype-based selection strategies require robust artificial epiphytotic conditions in the target ecologies which are resource-intensive and time consuming. Improved tropical and sub-tropical maize germplasm developed at CIMMYT in Mexico, are routinely deployed in SSA (especially at the maize breeding hubs of Kenya, Zimbabwe and Ethiopia) and vice versa. Adoption of maize varieties in SSA is mostly conditional upon reasonable levels of MSV resistance, along with high grain yield. Molecular markers that are associated with MSV resistance in a range of genetic backgrounds could potentially enable pre-selection of genomic regions in the tropical germplasm developed within and outside SSA, thereby contributing to enhanced and accelerated genetic gains.

A major QTL (*Msv1*) for MSV resistance on chromosome bin 1.05 has been reported in several studies on MSV resistance (Kyetere et al. 1999; Welz et al. 1998; Pernet et al. 1999a, b; Lagat et al. 2008; Babu et al. Manuscript in prep). The proportion of phenotypic variance explained by *Msv1* in various studies ranged from 24 to 92 %, and two reports identified the locus as a major gene (Kyetere et al. 1999; Welz et al. 1998). The physical confidence interval of this major effect QTL in various studies, with reference to the RefGen_V2 of B73, ranged from 16 Mb to 133 Mb. The large physical interval of this region on the short arm of chromosome 1 is possibly due to a number of factors including proximity to the centromere, low density marker coverage and limited number of recombinant individuals in the mapping populations of earlier studies. Despite the larger delimited interval, MAS for MSV resistance, using a SSR marker, *umc1917*, selected for its consistent polymorphism and co-dominance between the resistant and susceptible lines was found to be advantageous over conventional phenotype-based selection in terms of reduced disease incidence and improved resource efficiency (Abalo et al. 2009). Though transgenic lines with immunity to MSV have been developed using pathogen derived resistance (Shepherd et al. 2007), marker-assisted breeding for native trait enhancement is preferred over transgenic approaches because of ease of integration in the conventional breeding pipeline as well as regulatory, commercial and social acceptance barriers to transgenic adoption. Consequently, marker-assisted breeding for MSV resistance is one of the most pragmatic approaches to select for alleles of interest in elite breeding populations. Markers based on known functional polymorphisms within the target gene(s) are the first choice for MAS programs, however, such diagnostic candidate markers could be difficult and expensive to identify. Markers that are at a sub-centimorgan distance from the target gene/QTL are often the most economical proxies for enabling indirect selection for the trait. Nonetheless, a MAS approach of this nature is not always straight-forward in tropical maize germplasm, where high genetic diversity and rapid LD decay, can restrict applicability of marker–trait associations that are discovered in one particular background to other populations derived from either partially or unrelated germplasm. In the present study, LD decay occurred at the rate of approximately 0.7 kb in a tropical maize panel of 278 lines, which is much more rapid than LD decay rate in elite temperate maize germplasm, where LD decay was reported at about 500 kb (Rafalski 2002), and comparable to that observed in a set of diverse public breeding lines (Remington et al. 2001; Tenaillon et al. 2001). This necessitates identification of marker–trait associations that are in strong LD so as to enable selection over a range of diverse germplasm as well as many successive generations in the breeding programs. We have successfully fine-mapped the *Msv1* QTL to a 0.87 cM interval equivalent to a physical interval of 7 Mb. Three production marker assays, representing three SNPs in the most significant MSV-resistant haplotype, that are amenable for high-throughput automation and capable of accurately predicting the reaction to MSV infection in a diverse array of tropical germplasm have been developed.

QTL mapping of MSV resistance

Initial mapping for reaction to MSV based on 262 progenies of CML206 (resistant)/CML312 (susceptible) identified a large effect QTL on chromosome 1 in a 5.9 cM interval between PZA00944.1 and csu1138.4, which explained approximately 67 % of phenotypic variance. This QTL co-localised with *Msv1,* a large effect QTL identified in earlier studies involving multiple MSV-resistant lines and evaluated for reaction to MSV in various environments in Africa. *Msv1* was found to be partially dominant, which was in agreement with earlier studies (Welz et al. 1998; Pernet et al. 1999a, b). Due to this weak dominance effect, probably all parental components of a hybrid need to carry the favourable *Msv1* allele to reach a sufficient level of resistance in the hybrid. In this study, we have not attempted to validate this using test-crosses. In addition to *Msv1*, two other minor QTL were identified in the current study on chromosomes 3.05/06 and 10.04, which however did not largely overlap with previous studies that reported other loci for reaction to MSV infection, possibly due to environment and germplasm dependent nature of the minor effect QTL. Nevertheless, synergistic effect of these minor QTL with that of *Msv1* has been documented in an earlier study (Welz et al. 1998), which reported a QTL of smaller effect size on chromosome 3.04/06 which contributed greater levels of resistance in CML202 compared to another resistant line, Tzi4, in which resistance to MSV was largely conditioned by *Msv1*. This QTL interval overlaps with the interval of the minor QTL on chromosome 3 reported here. Although the favourable allele at *Msv1* locus appears to be shared by multiple resistant lines, rapid LD decay in tropical germplasm severely constraints the use of common flanking markers delimiting a broader interval. Identification of markers that are tightly linked to *Msv1* will enable selection in early generations in a range of populations, involving different resistant lines. This may enable undertaking field evaluation of breeding materials for their responses to MSV infection only at the later stages, thus helping to fix the minor effect loci that are relevant in specific genetic backgrounds/environments, in addition to *Msv1*.

Fine mapping of *Msv1*

Fine mapping of a large effect QTL in the presence of other minor effect loci requires some special considerations. We

designed a slight variant of the F_3-based QIR approach (Peleman et al. 2005), in which the genomic regions harbouring minor effect loci are selectively homogenised in one generation as against the NIL (near-isogenic line) approach, which seeks to homogenise the entire background except the target locus through repeated backcrosses. QIR strategy is time- and resource-efficient and is considered appropriate for oligogenic traits. In this strategy, after initial QTL mapping, informative plants (QIRs) are selected from a large F_2/F_3 population which carry a recombination event at the locus of interest and homogenised for the other non-target QTL, thereby reducing the complexity to almost that of a monogenic trait in an oligogenic population. We derived QIRs from an F_3 population, as opposed to F_2 in the original approach. This allowed us to fix the target locus and subsequently multiply seed for repeated phenotyping with enhanced precision in field evaluation of QIRs. The variant QIR approach adopted in this study also helped to control the partial dominance of *Msv1*, since heterozygosity at the locus in F_2-derived individuals could lead to ambiguous classification of recombinants.

Seed DNA-based genotyping (Gao et al. 2008) enabled non-destructive and efficient screening of a large number (4725) of F_2 seeds for the flanking markers of the *Msv1* interval. Additional markers placed outside the flanking markers on either side were included in order to avoid the risk of missing the target locus due to the relative imprecision of the QTL interval based on a map derived from a smaller number of progenies. The phenotypes of 36 QIRs were compared to control sets and categorised into genotypically resistant and susceptible classes for the *Msv1* interval. Identification of control QIR sets with both resistant and susceptible background QTL effects allowed delineation of the effect of *Msv1* from the effect of the alleles at other QTL. The *Msv1* interval was saturated with additional markers and seven recombination exchange points were identified, enabling delimitation of *Msv1* to a smaller region of 7 Mb/0.87 cM. Recombination was generally lower in this region, possibly due to its peri-centromeric positioning. Interestingly, the fine-mapped interval was slightly outside the confidence interval of the originally identified QTL, which could be due to the inaccurate localization based on a limited set of 253 $F_{2:3}$ families. The refined *Msv1* interval was resolved to the sub-cM level, but still represented ~7 Mb of the physical distance. Hence, the chances are minimal that a subsequent effort with an additional increase in population size utilising the same parental cross could appreciably dissect this region further. MAS based on flanking markers that cover a large physical interval of ~7 Mb may not be very efficient in selecting for *Msv1*, especially when used in partially related or unrelated populations derived from wider germplasm pools. Also, the large physical distance could translate into unwanted linkage drag in such a recombination-suppressed region. As an alternative to further fine mapping using larger number of individuals, we sought to explore the high-density genotype information generated using GBS in the DTMA-AM panel which was phenotyped against MSV in order to validate the fine-mapped interval and to further resolve the interval for use in a broader germplasm context.

GWAS and discovery of production markers for *Msv1*

GWAS maps QTL using an open system design where relatedness within the study population is unknown, as opposed to controlled biparental populations that are artificially created (Myles et al. 2009). Both approaches have merits and demerits (comprehensive reviews of the subject in Nordborg and Weigel 2008; Zhu et al. 2008; Ersoz et al. 2009; Myles et al. 2009; Rafalski 2010), particularly with regard to the power and mapping resolution. A particular advantage of GWAS over biparental mapping is the applicability of resulting marker–trait associations over broader set of diverse individuals as compared to population specific inference of biparental mapping (Zhu et al. 2008). The association mapping panel used in this study (DTMA-AM) broadly represents tropical and subtropical maize genetic diversity, and has been widely used at CIMMYT for identifying marker–trait associations for an array of traits for maize in the tropics. This panel was genotyped using GBS platform, which currently generates close to one million SNPs based on a global repository of ~30,000 maize genotypes available at the Institute of Genomic Diversity, Cornell University (Romay et al. 2013). Mixed model analyses effective in correcting for population structure and relatedness in the association mapping panel (Yu et al. 2008) was employed in the present study. We identified 19 significant SNPs between 82 and 93 Mb on chromosome 1, of which S1_87301459 (87.3 Mb) was found to be the most significantly associated with MSV resistance. The favourable allele at this locus was found in ~30 % of the lines in the DTMA-AM panel with an allele substitution effect of 1.33 units. This GWAS-identified *Msv1* interval overlapped with the previously delimited interval based on the biparental population (described earlier). The most significant SNP was located in the 10th exon of the gene, GRMZM2G046848 (Fig. 3), which is a U-box domain-containing tyrosine kinase family protein (www.gramene.org). Additionally, two more SNPs located within this gene were found to be significantly associated with the trait. A number of protein kinases have been established as disease resistance genes (reviewed in Bent, 1996) and U-box domains that are part of E3 ligases have been demonstrated to be associated with plant defense mechanism through ubiquitination-mediated protein degradation (Zeng et al. 2006). However, further experiments will be needed to

elucidate the functional significance of this gene in relation to MSV resistance, if any. Based on the previous QTL mapping studies, the *Msv1* locus was placed between 64.9 and 67.4 Mb on Chromosome 1 based on nearest markers (TIDP5725 and IDP2553) in the IBM2 2008 Neighbours map (http://www.maizegdb.org). It is to be noted that *Msv1* has not yet been physically mapped and the above coordinates are an extrapolation from earlier QTL mapping studies. Our study has empirically delimited the *Msv1* locus to the physical interval between 82.7 and 89.2 Mb on chromosome 1, and the physical interval of 65–73 Mb was eliminated by the recombination break point analysis in the large mapping population we studied (Fig. 2). Moreover, this delimited interval is well supported by the genome-wide association analysis in an independent AM panel.

The LD decay in the DTMA-AM panel was rapid, with an average distance of about 690 bp ($r^2 = 0.1$), which is comparable to that reported by Romay et al. (2013) within 1 Kb in tropical germplasm, whereas the temperate subgroups in the same study showed a much slower decay. The comparative LD analysis between *Msv1* region and the whole genome revealed slower LD decay in the delimited *Msv1* interval (82–93 Mb), which could favour marker-assisted selection in a relatively wider germplasm due to extended marker-trait associations. However, such a slower LD decay may not allow further resolution of the locus unless a larger number of lines representing broader germplasm collections are included and characterised in the association panel. Individual SNPs are not as informative as diagnostic markers, as the expected heterozygosity is approximately three times smaller than that of SSR markers. However, if the expected heterozygosity is calculated on the basis of haplotypes rather than individual SNPs, the value is twice higher (Ching et al. 2002). A set of 2–4 SNPs is considered ideal for defining haplotypes on the basis of both informativeness and cost-effectiveness (Johnson et al. 2001). We detected 7 haplotype blocks from the 19 SNPs, of which Haplo_5, consisting of three SNPs, S1_87301345 (PZE-101093951), S1_ 87301459 (PZE0186065237) and PZE0186365075 spanning a 300 Kb region was found to be most significantly associated with the MSV resistance. We further validated the predictability of this haplotype block in an independent set of elite tropical/subtropical breeding lines, including CIMMYT maize lines (CMLs), with extreme MSV phenotypes. Of the 7 haplotype alleles identified among these lines, Haplo_5.1 was associated with resistance while Haplo_5.2 with susceptibility in almost all their occurrences. Our study found that the Haplo_5.2 allele was highly predictive of MSV susceptibility, with a co-occurrence of 94 %. The association of the identified Haplo_5 block with the MSV reaction was further validated additionally in two biparental populations between MSV-resistant and susceptible lines. In both the cases, the resistant haplotype was found to have a significant effect on reducing the mean disease by 1.03–1.39 units on a 1–5 scale. High-throughput single-plex SNP assays (KASP) have been developed for the three markers defining the haplotype block. KASP (Semagn et al. 2012) is fast becoming a global benchmark for low-cost genotyping and offers resource efficiency and flexibility at scale, especially in a "many-individuals-few markers" context, fitting the requirement of screening for MSV resistance in most of the maize breeding programs in the SSA.

Application of *Msv1* production markers in maize breeding in Africa

Undoubtedly, MSV is one of the major biotic stresses that the maize breeding programs in SSA routinely screen for. Climate change is likely to impact SSA in a significant manner and consequently intensive efforts are being made by institutions like CIMMYT in SSA to develop and deploy climate-resilient germplasm with a range of abiotic and biotic stress tolerance (Cairns et al. 2013). Tolerance to multiple stresses such as drought, heat and poor soil fertility along with resistance to a number of diseases such as MSV, maize lethal necrosis (MLN), NCLB and GLS have become imperative for sustainably managing maize production in SSA (Prasanna et al. 2014), especially in the hotspots of climate-vulnerability. CIMMYT and IITA, in partnership with several institutions in SSA, have developed an array of donor lines with resistance to various abiotic and biotic stresses; this germplasm is very widely distributed and used by both public and private sector institutions. Breeders' efforts to develop germplasm with tolerance/resistance to multiple stress tolerance in the SSA often involves crossing MSV-resistant lines with other breeding materials that may not have resistance to MSV (although these may have other adaptive traits), and hence, require an efficient trait-tracking mechanism during the advancement of segregating generations. The three marker haplotype reported here can efficiently predict reaction to MSV infection across a range of genetic backgrounds in a cost-effective manner. *Msv1* has been reported in almost all MSV-resistant lines identified so far and is generally considered an essential prerequisite for reasonable levels of MSV resistance. In order to avoid the over-dependence on *Msv1*, marker-based screening for *Msv1* in early generations (F_2, F_3, BC_1) may also be followed up with successive seasons of late generation field phenotyping for reaction to MSV infection, thereby concentrating the minor effect loci that could act synergistically with *Msv1* in conferring enhanced and durable resistance. Enhanced resistance through additional phenotypic selection will also help prevent possible breakdown of *Msv1*-mediated resistance in the long term, as multiple strains of MSV-A are reported to sporadically co-occur with the most virulent strain, MSV-A1 in different parts of SSA (Shepherd et al.

2010). Use of markers will significantly expedite genetic gain for MSV resistance, especially when selection for other traits such as drought tolerance is the major consideration. The *Msv1*-associated markers reported in this study are already being routinely employed by CIMMYT maize breeding programs in the development of superior germplasm targeted for deployment in the SSA, including selection/discarding of early generation progenies, and subsequently subjecting the selected source populations for haploid induction and DH line development.

Author contribution statement Designed the experiments: SN, RB; Conducted the experiments: SN, RB, CM, GM, KS, YB, BD, DM, LK; Analysed the data: SN, RB; Contributed materials/analysis tools: SN, RB; Wrote the manuscript: SN, RB, MO, BMP.

Acknowledgments The authors gratefully acknowledge the financial support received from the Bill and Melinda Gates Foundation (BMGF) as part of the project, "Drought Tolerant Maize for Africa (DTMA)". We thank CGIAR Research Program (CRP) on MAIZE for co-sponsoring this research work. The biparental populations and disease data from Kevin Pixley, CIMMYT and Jean-Marcel Ribaut, GCP, used in initial QTL mapping and validation study are thankfully acknowledged. Assistance for data analysis provided by Jyothsna Tejomurthula is appreciated. The authors would also like to thank the technical assistance from Carlos Martinez, Alberto Vergara and Jose Simon Marias of CIMMYT in carrying out this research work. The authors also gratefully acknowledge the two reviewers for their valuable comments that could improve the value of this manuscript.

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standards The authors declare no ethical standards have been violated in the course of the study.

References

Abalo G, Tongoona P, Derera J, Edema R (2009) A comparative analysis of conventional and marker-assisted selection methods in breeding maize streak virus resistance in maize. Crop Sci 49:509–520

Bent AF (1996) Plant disease resistance genes: function meets structure. Plant Cell 8:1757–1771

Bishop DT, Williamson JA (1990) The power of identity-by-state methods for linkage analysis. Am J Hum Genet 46:254–265

Cairns JE, Crossa J, Zaidi PH et al (2013) Identification of drought, heat, and combined drought and heat tolerant donors in maize. Crop Sci 53:1335–1346

Ching A, Caldwell KS, Jung M et al (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19

CIMMYT (2001) Laboratory protocols: CIMMYT applied molecular genetics laboratory protocols. CIMMYT, Mexico

Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6(5):e19379

Ersoz E, Yu J, Buckler E (2009) Applications of linkage disequilibrium and association mapping in maize. In: Kirz A, Larkins B (eds) Molecular genetic approaches to maize improvement. Springer, Berlin, pp 173–195

Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 12:921–927

Gabriel SB, Schaffner SF, Nguyen H et al (2002) The structure of haplotype blocks in the human genome. Science 296:2225–2229

Gao S, Martinez C, Skinner DJ et al (2008) Development of a seed DNA-based genotyping system for marker-assisted selection in maize. Mol Breed 22:477–494

Guthrie EJ (1978) Measurement of yield losses caused by maize streak disease. Plant Dis Rep 62:839–840

Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol 33:54–78

Johnson GCL, Esposito L, Barratt BJ et al (2001) Haplotype tagging for the identification of common disease genes. Nat Genet 29:233–237

Kim SK, Efron Y, Khadr F, Fajemisin J, Lee MH (1987) Registration of 16 maize streak resistant tropical maize parental inbred lines. Crop Sci 27:824–825

Kim SK, Efron Y, Fajemisin JM, Buddenhagen IW (1989) Mode of gene action for resistance in maize to maize streak virus. Crop Sci 29:890–894

Kyetere DT, Ming R, McMullen MD et al (1999) Genetic analysis of tolerance to maize streak virus in maize. Genome 42:20–26

Lagat M, Danson M, Kimani M, Kuria A (2008) Quantitative trait loci for resistance to maize streak virus in maize genotypes used in hybrid development. Afric J Biotech 7:2573–2577

Martin DP, Shepherd DN (2009) The epidemiology, economic impact and control of maize streak disease. Food Sec 1:305–315

Martin DP, Willment JA, Billharz R et al (2001) Sequence diversity and virulence in Zea mays of Maize streak virus isolates. Virology 288:247–255

Myles S, Peiffer J, Brown P et al (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21:2194–2202

Nordborg M, Weigel D (2008) Next-generation genetics in plants. Nature 456:720–723

Peleman JD, Wye C, Zethof J et al (2005) Quantitative trait locus (QTL) isogenic recombinant analysis: a method for high resolution mapping of QTL within a single population. Genetics 171:1341–1352

Pernet A, Hoisington D, Franco J et al (1999a) Genetic mapping of maize streak virus resistance from the Mascarene source. I. Resistance in line D211 and stability against different virus clones. Theor Appl Genet 99:524–539

Pernet A, Hoisington D, Dintinger J et al (1999b) Genetic mapping of maize streak virus resistance from the Mascarene source. II. Resistance in line CIRAD390 and stability across germplasm. Theor Appl Genet 99:540–553

Prasanna BM, Chaikam V, Mahuku G (2012) Doubled haploid (DH) technology in maize breeding: an overview. In: Prasanna BM, Chaikam V, Mahuku G (eds) Doubled haploid technology in maize breeding: theory and practice. CIMMYT, Mexico, pp 1–8

Prasanna BM, Babu R, Nair S et al (2014) Molecular marker-assisted breeding for tropical maize improvement. In: Wusurika R, Bohn M, Lai J, Kole C (eds) Genetics, genomics and breeding of maize. CRC Press, London, pp 89–119

Pratt RC, Gordon SG (2006) Breeding for resistance to maize foliar pathogens. Plant Breed Rev 27:119–174

R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Rafalski JA (2010) Association genetics in crop improvement. Curr Opin Plant Biol 13:174–180

Remington DL, Thornsberry JM, Matsuoka Y et al (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Nat Acad Sci USA 98:11479–11484

Rodier A, Assie J, Marchand JL, Herve Y (1995) Breeding maize lines for complete and partial resistance to maize streak virus (MSV). Euphytica 81:57–70

Romay MC, Millard MJ, Glaubitz JC et al (2013) Comprehensive genotyping of the USA national maize inbred seeds bank. Genome Biol 14:R55

Semagn K, Beyene Y, Makumbi D et al (2012) Quality control genotyping for assessment of genetic identity and purity in diverse tropical maize inbred lines. Theor Appl Genet 125:1487–1501

Shepherd DN, Mangwende T, Martin DP et al (2007) Maize streak virus-resistant transgenic maize: a first for Africa. Plant Biotech J 5:759–767

Shepherd DN, Martin DP, Van der walt E et al (2010) Maize streak virus: an old and complex 'emerging' pathogen. Mol Plant Path 11:1–12

Soto PE, Buddenhagen IW, Asnani VL (1982) Development of streak virus-resistant maize populations through improved challenge and selection methods. Ann Appl Biol 100:539–546

Storey HH, Howland AK (1967) Inheritance of resistance in maize to the virus of streak disease in East Africa. Ann Appl Biol 59:429–436

Stuber CW, Lincoln SF, Wolff DW (1987) Molecular markers facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits. Crop Sci 27:639–648

Tang CY, Bjarnason MJ (1993) Two approaches for the development of maize germplasm resistant to maize streak virus. Maydica 38:301–307

Tenaillon MI, Sawkins MC, Long AD et al (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.). Proc Natl Acad Sci USA 98:9161–9166

Wambugu F, Wafula J (2000) Advances in maize streak virus disease research in Eastern and Southern Africa. Workshop Report, 15–17 September 1999, KARI and ISAAA AfriCenter, Nairobi, Kenya. ISAAA Briefs No. 16. ISAAA, Ithaca, NY, p 43

Welz HG, Schechert A, Pernet A et al (1998) A gene for resistance to maize streak virus in the African CIMMYT maize inbred CML202. Mol Breed 4:147–154

Yan J, Shah T, Warburton M et al (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS One 04:e8451

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178:539–551

Zeng LR, Vega-Sanchez ME, Zhu T, Wang GL (2006) Ubiquitination-mediated protein degradation and modification: an emerging theme in plant-microbe interactions. Cell Res 16:413–426

Zhu C, Gore M, Buckler E, Yu J (2008) Status and prospects of association mapping in plants. Plant Genome 1:5–20