ORIGINAL PAPER

Prospecting sugarcane resistance to Sugarcane yellow leaf virus by genome-wide association

S. Debibakas · S. Rocher · O. Garsmeur · L. Toubi ·
D. Roques · A. D'Hont · J.-Y. Hoarau · J. H. Daugrois

Abstract

Key message **Using GWAS approaches, we detected independent resistant markers in sugarcane towards a vectored virus disease. Based on comparative genomics, several candidate genes potentially involved in virus/aphid/plant interactions were pinpointed.**

Abstract Yellow leaf of sugarcane is an emerging viral disease whose causal agent is a Polerovirus, the Sugarcane yellow leaf virus (SCYLV) transmitted by aphids. To identify quantitative trait loci controlling resistance to yellow leaf which are of direct relevance for breeding, we undertook a genome-wide association study (GWAS) on a sugarcane cultivar panel ($n = 189$) representative of current breeding germplasm. This panel was fingerprinted with 3,949 polymorphic markers (DArT and AFLP). The panel was phenotyped for SCYLV infection in leaves and stalks in two trials for two crop cycles, under natural disease pressure prevalent in Guadeloupe. Mixed linear models including co-factors representing population structure fixed effects and pairwise kinship random effects provided an efficient control of the risk of inflated type-I error at a genome-wide level. Six independent markers were significantly detected in association with SCYLV resistance phenotype. These markers explained individually between 9 and 14 % of the disease variation of the cultivar panel. Their frequency in the panel was relatively low (8–20 %). Among them, two markers were detected repeatedly across the GWAS exercises based on the different disease resistance parameters. These two markers could be blasted on *Sorghum bicolor* genome and candidate genes potentially involved in plant–aphid or plant–virus interactions were localized in the vicinity of sorghum homologs of sugarcane markers. Our results illustrate the potential of GWAS approaches to prospect among sugarcane germplasm for accessions likely bearing resistance alleles of significant effect useful in breeding programs.

S. Debibakas
Université des Antilles et de la Guyane, 97157 Pointe-à-Pitre, Guadeloupe, France

S. Debibakas · J. H. Daugrois
CIRAD, UMR BGPI, 97170 Petit Bourg, Guadeloupe, France

S. Rocher · O. Garsmeur · A. D'Hont
CIRAD, UMR AGAP, 34398 Montpellier, France

S. Rocher · L. Toubi · D. Roques · J.-Y. Hoarau (✉)
CIRAD, UMR AGAP, 97170 Petit Bourg, Guadeloupe, France
e-mail: jean-yves.hoarau@cirad.fr

Introduction

Yellow leaf disease of sugarcane (*Saccharum* spp.) is an important and widely spread disease causing severe yield losses, ranging from 20 to 40 % in susceptible cultivars (Lockhart and Cronjé 2000; Zhu et al. 2011; Rassaby et al. 2003; Vega et al. 1997). This disease has been well described (Schenck 2001; Rott et al. 2008) and is caused by the Sugarcane yellow leaf virus (SCYLV), a polerovirus (D'Arcy and Domier 2005) vectored in a persistent, circulative, non-propagative manner by different aphids species as *Ceratovacuna lanigera, Melanaphis sacchari, Rhopalosiphum maidis*, and *R. rufiabdominalis* (Schenck and Lehrer 2000; Zhou et al. 2006). *M. sacchari* has been shown to be the most efficient aphid vector of SCYLV

when compared with *R. maidis*, and *R. rufiabdominalis* (Schenck and Lehrer 2000).

Yellow leaf spread by aphids depends on cultivar susceptibility and local epidemiological conditions as well as climatic parameters (Daugrois et al. 2011) and aphid predator populations. The disease is also spread by the use of infected cuttings as seed cane in planting operations. Currently, little is known on the dynamics of the virus within infected sugarcane (circulation and maintenance) with respect to the complex anatomy stalks which emerge from partially persistent root systems across successive crop cycles. In addition, because of asymptomatic and non-specific foliar symptoms (yellowing of leaf midrib), the spread of the disease may be insidious. The only reliable and rapid methods to confirm the presences of the virus are by immunoassay (Schenck et al. 1997) or PCR approaches (Comstock et al. 1998).

SCYLV was first identified in Guadeloupe in 1996 (Daugrois et al. 1999) where the virus is vectored by the aphid *M. sacchari* (Zehntner). This virus is present in all commercial sugarcane cultivars and in all sugarcane growing areas in Guadeloupe, where field SCYLV incidences varied from 0 to 21 % in 2005 depending on cultivars and locations (Edon-Jock et al. 2007). Incidences have since increased to reach, in 2010, between 1 and 72 % for the most susceptible cultivars (personal communication, J.H. Daugrois).

Varietal resistance improvement is the most efficient strategy to control sugarcane diseases, but the lack of knowledge of the genetic basis of yellow leaf resistance makes breeding progress for this disease difficult. Some studies have dealt with the exploration of genotypic variability of resistance to SCYLV in sugarcane germplasm. Indexation of germplasm collections of sugarcane and related species for the presence of the virus revealed potential sources of resistance to yellow leaf disease. Commercial cultivars (*Saccharum* spp.), *S. officinarum*, *S. robustum* and *S. sinense* presented high virus incidence, whereas virus incidence was low for *S. barberi and S. spontaneum*, and the related genus *Miscanthus* spp. and *Erianthus* spp. (Comstock et al. 2001; Komor 2011). Until today, only one genetic study was attempted to characterize resistance to yellow leaf disease. Using a QTL approach based on a biparental progeny between a susceptible cultivar and a resistant clone, Costet et al. (2012) tagged the first major quantitative trait allele (QTA) of resistance to SCLYV (named *Ryl*1). Identifying additional QTLs for resistance would allow for greater flexibility in breeding programs, as well as allowing resistance pyramid strategies to be employed. To this end, a survey of the current sugarcane germplasm in a genome-wide association mapping represents an attractive approach to search for additional resistance genes.

The possibility of applying linkage disequilibrium (LD) based studies, or association mapping studies, to identify marker-trait associations in sugarcane has been highlighted earlier given the relatively large extent of LD existing among modern cultivars (Jannoo et al. 1999). The large extent of LD is attributed to the recent breeding history characterized by only few generations separating modern cultivars from a limited number of founder clones. LD is strong in the first five centimorgans and generally drops sharply when markers are 5 cM or more apart, although instances of LD extending up to 10 or 20 cM are not rare (Raboin et al. 2008). Modern sugarcane cultivars (*Saccharum* spp.) derive from introgressions into the highly polyploid domesticated sugar-producing species *S. officinarum*, ($x = 10$ and $2n = 8x = 80$) of the wild *S. spontaneum* species characterized by different cytotype levels ($x = 8$ and $2n = 5x - 16x = 40 - 128$) (D'Hont et al. 1998; Sreenivasan et al. 1987). As a result, sugarcane cultivars have a large and complex genome of about 120 chromosomes (10 Gb) corresponding to about 12 homologous sets of a monoploid genome of 1,500 cM. Notwithstanding genome complexity, several studies have applied association mapping to sugarcane. Using 1,209 polymorphic markers (AFLP and SSR), Wei et al. (2006) detected associations with resistance to four diseases in a collection of 154 cultivars and later on they identified DArT markers related to cane yield and sugar content traits on a sample of 480 genotypes (Wei et al. 2010). These studies confirmed the potential of genome-wide association mapping in the polyploid context of sugarcane.

The purpose of this work was to prospect for sugarcane resistances to SCYLV disease using a genome-wide association approach to identify markers linked to genomic regions involved in the control of the disease.

Materials and methods

Plant material

A panel of 189 sugarcane accessions originated from 26 sugarcane breeding programs around the world (supplementary material 1) was planted in two successive trials (A and B) at Station de Roujol, CIRAD, Petit Bourg, Guadeloupe, French West Indies. This panel was composed of a large majority of advanced interspecific sugar-producing commercial hybrids and a very few early-generation interspecific hybrids. Both trials were planted in a randomized complete block design with three complete blocks and a 4 m sugarcane row as an experimental plot. Plots were separated by 1.5 m between line and rows. Trials were conducted during three successive crop cycles. Both trials were planted using cuttings from the germplasm collection of

CIRAD Petit Bourg which had been maintained locally for at least 5 years under natural infection. Trial A was planted on 28 September, 2005. Dates of successive crop harvest were 22 May 2006, 26 March 2007 and 17 March, 2008 for plant cane (PC), first ratoon (R1) and second ratoon (R2) crops. Trial B was planted on 26 September, 2007. Dates of successive crop harvest were 11 June 2008, 28 May 2009 and 15 March 2010 for PC, R1 and R2 crop. Due to the inter-annual variation in weather and crop conditions, trials A and B underwent different vector pressures. Population of aphids was abundant in plant cane in trial A, as evidenced by sooty mold that accompanies severe infestations of the aphid in sugarcane (Hall and Bennett 1994) and absence of other sap feeding insects, whereas aphids were sparsely observed in ratoon crops of trial A and in all crop cycles of trial B where no sooty mold was observed.

Sampling for SCYLV diagnoses by tissue blot immunoassay (TBIA)

In both trials (A and B), assessments of the cultivar panel for yellow leaf were achieved using diagnoses of leaf and stalk infection using tissue blot immunoassay (TBIA). A total of 10–15 first visible dewlap (FVD) leaves were sampled per experimental plot on 5-month-old plants in PC and R2 crop years. Leaf imprints were made by printing a cross section of the bottom part of the midrib of the FVD leaf. In addition, six (PC trial A) and ten (PC trial B and R2 of both trials) stalks were randomly sampled per plot before harvest date. A longitudinal core of 1 cm diameter was taken from an internode of the one-third lower part of the stalk. A stalk imprint was done with a transversal section of the core as done for ratoon stunting disease diagnosis (Davis et al. 1994).

Leaf and stalk imprints were made by applying hand-pressure to the transversal section of plant tissues on a nitrocellulose membrane (Millipore, Protran BA85, 0, 45 μm) for one second.

SCYLV serological diagnosis of tissue imprints

SCYLV in leaf and stalk samples was diagnosed by following the method described by Schenck et al. (1997) modified by Daugrois et al. (2011). Briefly after being blocked with BSA, the membranes were incubated in anti-SCYLV AS-R2 IgG (kindly provided by Pr B.E.L. Lockhart). The membranes were washed in TBS-Tween (0.05 %), and incubated in anti-rabbit IgG alkaline phosphatase conjugate antibody produced in goat (Sigma-Aldrich A3812). After washing, the membranes were incubated in a substrate solution NBT/BCIP tablets (Sigma-Aldrich B5655). Positive samples and infection intensity of imprints were read on wetted processed membranes under a diascopy binocular microscope.

Disease resistance traits

For each experimental plot, we estimated the virus incidence in leaves and stalks based on the percentage of positive leaves and stalks, respectively. We also estimated the virus population density within both plant tissues by computing for each plot the means of the infection intensity observed on imprints on the basis of a 0–5 scale. For leaf imprints, this intensity scale was: 0: no vascular bundle infected; 1: infected vascular bundles (ivb) <10 %; 2: 10 % ≤ ivb <50 %; 3: 50 % ≤ ivb <100 %; 4: 100 % ivb with slight coloration; and 5: 100 % ivb with intense coloration. For stalk imprints, the scale was modified as follows: 1: 0 % < ivb <5 %; 2: 5 % ≤ ivb <30 %; 3: 30 % ≤ ivb <60 %; 4: 60 % ≤ ivb <100 %; and 5:100 % ivb.

Statistical analyses of phenotypic data

All statistical analyses were conducted using the SAS software (SAS software 9.3, SAS Institute Inc., Cary, NC, USA). Virus incidence (VI) data were transformed by the $ar\sin\left(\sqrt{VI}\right)$ function before statistical analyses to achieve normal distribution of residuals. Independency of cultivar means and variances were controlled for transformed virus incidence data and for virus population density data.

Analyses of variance (ANOVA) were done for each disease resistance trait considering mixed linear models (SAS Mixed procedure) with the restricted maximum likelihood (REML) method. Each trait was analyzed with the two following models considering either both trails together (model 1) or each trial analyzed separately (model 2):Model 1:

$$\underline{Y}_{ijkl} = \mu + \underline{G}_i + B_{j(l)} + C_{k(l)} + T_l + \underline{GC}_{ik(l)} + \underline{GT}_{il} + \underline{\varepsilon}_{ijkl}$$

Model 2:

$$\underline{Y}_{ijk} = \mu + \underline{G}_i + B_j + C_k + \underline{GC}_{ik} + \underline{\varepsilon}_{ijk}$$

In both models, terms underlined indicate random effects, while other effects are fixed. $\underline{Y}_{ijk}$ and $\underline{Y}_{ijkl}$ are the phenotype of the jth block ($j = 1, 2, 3$) of the ith genotype ($i = 1, 2, …, 189$) of kth crop cycle ($k = 1, 3$) in each trial or in the lth trail ($l = 1,2$); μ is the overall mean; $\underline{G}_i$ is the effect of individual i; $B_{j(l)}$ and B_j are block effect j nested or not nested within trial l, respectively; $C_{k(l)}$ and C_k are crop cycle effect k nested or not nested within trial l; T_l is the trial effect; $\underline{GC}_{ik(l)}$ and $\underline{GC}_{ik}$ are the genotype x crop cycle interaction nested or not nested within trial l; $\underline{GT}_{il}$ is the genotype × trial interaction; $\underline{\varepsilon}_{ijkl}$ and $\underline{\varepsilon}_{ijk}$ are residual errors. In both models individual plot phenotypes within a trial were considered as repeated measurements across crop cycles. The significance of each variance component was tested by a Wald Z test using SAS Mixed procedure. The

significance of the differences between means of traits per crop cycle within and between trials, were tested with a *t* test under Lsmeans statement in the analysis of model 1 under SAS Mixed procedure.

Broad-sense heritability (H^2) at an experimental level was evaluated for each trait as the ratio of genotypic variance to phenotypic variance, using the components of variance obtained from the ANOVA as follows:

$$H^2 = \sigma_G^2/(\sigma_G^2 + \sigma_{GC}^2/C + \sigma_{GT}^2/T + \sigma_e^2/TCR) \text{ for model 1}$$

$$H^2 = \sigma_G^2/(\sigma_G^2 + \sigma_{GC}^2/C + \sigma_e^2/CR) \text{ for model 2}$$

where σ_G^2 = estimation of the genotypic variance; σ_{GC}^2 = estimation of the variance of genotype × crop cycle interaction effect; σ_{GT}^2 = estimation of genotype × trial interaction; σ_e^2 = estimation of the residual variance; C = number of crop cycles (2); T = number of trial (2) and R = number of replications (3).

Phenotypic cultivar values within each model were estimated by adding to their empirical best linear unbiased predictors (EBLUPs), the intercept estimate and all fixed effects estimates. Phenotypic values (PhV) were then used to correlate disease resistance traits among and within trials by calculating Pearson correlation coefficients and to perform marker-trait association tests.

Genotyping data

The Amplified Fragment-Length Polymorphism (AFLP) genotyping was generated using the AFLP® Analysis System I (Invitrogen), with 59 primer pairs, as recommended by the manufacturer with slight modifications described by Hoarau et al. (2001). The Diversity Arrays Technology (DArT) genotyping was produced by the commercial company Diversity Arrays Technology Pty Ltd. (Yarralumla, Australia) using their commercial sugarcane array (Heller-Uszynska et al. 2011). A total of 3,949 polymorphic markers exhibiting frequencies higher than 0.05 and lower than 0.95 (2,400 AFLP + 1,549 DArT) were used for the association mapping tests.

Marker-trait associations

These tests were conducted using TASSEL (online version 3.0.154) software with genotypic cultivar values either obtained from both trials together (model 1) or from each trail separately (model 2). Two association models were used: the General Linear Model (GLM) with a Q matrix indicative of population structure and the Mixed Linear Model (MLM) with both population structure (Q) and kinship (K) matrix as covariates (Yu et al. 2006). MLM analysis used the compressed mixed linear model approach (Zhang et al. 2010) carried by TASSEL.

Haplotypes were constructed by testing the 3,949 available polymorphic markers in pairwise associations with a bilateral Fisher exact test and by grouping all significant pairwise combinations of markers by transitivity (Raboin et al. 2008). To achieve an overall significance of 5 % using the Bonferroni correction, a nominal significance threshold of $P = 6.4 \times 10^{-9}$ was applied to each of the 7,795,326 pairwise tests so far performed ($3,949 \times 3,948/2$). We therefore found a total of 463 haplotypes (encompassing 2,480 markers) and 1,469 independent markers.

Q matrix was inferred from DArT markers using the Principal Component Analysis (PCA) approach developed by Patterson et al. (2006) with SAS Princomp procedure. The PCA analyses were performed on the cultivar panel by considering the only 1,509 polymorphic DArT markers showing <10 % of missing data among genotyped cultivars and exhibiting frequencies situated in a (0.05–0.95) interval. These DArT markers represented 619 different haplotypes and independent markers. However, the study of a relatively modest number of markers to cover the large genome of sugarcane (about 120 chromosomes) combined with the rather large linkage disequilibrium (LD) existing among sugarcane cultivars may result in the observation of markers in significant LD that could be separated by large genetic distances (Raboin et al. 2008). In this context, two alternative Q1 and a Q2 matrix were inferred from a PCA analysis either based on the 619 independent DArT haplotype/markers or on all the 1,509 DArT markers, respectively. The effect of these two alternative Q matrixes on GLM model (Q1 versus Q2) and on MLM model (Q1 + K versus Q2 + K) was compared based on the shape of Quantile–Quantile plot of the probabilities of corresponding association tests. The number of PCA axes used to define the Q matrix used in MLM and GLM models was determined according to Zhu and Yu (2009) methodology. Cultivar phenotypes (PhV) were fitted in a linear model with an increasing number of PCA axes and we chose the first most significant axes showing the lowest Bayesian Information Criteria (BIC). In both GLM and MLM models, the Q matrix was considered as a fixed co-variable. In the compressed MLM model, the additional kinship matrix (K) was a Jaccard double similarity matrix computed with Darwin software (Perrier and Jacquemoud-Collet 2006) considering the same 1,509 markers set used for the inference of Q matrix.

Inflation of false-positive associations may occur in genome-wide association studies when the genetic structure is not well modeled. A widely used inflation factor λ can be computed by dividing the median test statistic of the quantile–quantile plot by its theoretical median statistic under the null hypothesis of no marker linked to the polymorphism controlling the phenotypes (Price et al.

Table 1 Estimates and significance level of variance components and heritability (H^2) for SCYLV disease resistance traits

Trait	Model	Trial	Variance components				H^2
			σ_G^2	σ_{GC}^2	σ_{GT}^2	σ_e^2	
Leaf SCYLV incidence	1	A and B	0.176****	0.012****	0.049****	0.056****	0.83
	2	A	0.164****	0.017****		0.038****	0.92
	2	B	0.288****	0.008***		0.073****	0.95
Stalk SCYLV incidence	1	A and B	0.112****	0.011***	0.067****	0.102****	0.70
	2	A	0.117****	0.016****		0.093****	0.83
	2	B	0.243****	0.005 NS		0.111****	0.92
Leaf SCYLV density	1	A and B	1.130****	0.122****	0.236****	0.453****	0.84
	2	A	1.251****	0.179****		0.417****	0.89
	2	B	1.492****	0.064***		0.488****	0.93
Stalk SCYLV density	1	A and B	0.469****	0.074****	0.182****	0.380****	0.75
	2	A	0.483****	0.128****		0.354****	0.80
	2	B	0.822****	0.020 NS		0.405****	0.91

, * Significant for a Wald Z test at $P = 0.001$, $P = 0.0001$ or non-significant (NS) at $P = 0.05$

2010). In our study, the inflation factor λ was computed from the Fisher F-statistics provided by association models, following Yu et al. (2006). According to Price et al. (2010) λ should be lower than 1.05 to avoid detection of spurious associations. When λ ≈ 1, there is no inflation in test statistics.

For both GLM and MLM models with a controlled inflation risk (λ), associations were declared significant up to a genome wise type-I error rate (GWER) threshold of $P_{GWER} = 10$ %. For the GLM model, this threshold was defined on the basis of 1,000 permutations. For MLM model, considering the number of independent haplotypes or markers (1,932) this P_{GWER} of 10 % corresponded to $P = (0.10/1,932)$ using the Bonferroni correction and gives a $-\log P = 4.286$.

Finally, in each trial, for each disease resistance trait, we estimated the proportion of the total phenotypic variation that could be explained by the global effect (R^2) of all the detected markers using a multiple regression analysis (SAS REG procedure) with a stepwise process adding markers one-by-one. Entry and stay option parameters were set up for F test marker effect on the regression at $P = 0.15$.

Cumulative markers effect on SCYLV resistance was analyzed by comparing trait means of cultivar classes considering number of markers borne by cultivars, by a t test using Lsmeans statement under GLM procedure. Data used were the cultivar PhV means from model 1.

Search for genes involved in disease resistance

Since sugarcane and sorghum display high micro-colinearity (Jannoo et al. 2007; Wang et al. 2010; Garsmeur et al. 2010), the sequences from selected sugarcane DArT markers were blasted on *Sorghum bicolor* sequence genome using BLASTN 2.2.22 program (Altschul et al. 1997). Genes surrounding sugarcane sequences were searched on sorghum sequence using a window of 100 kb on each side using the Gbrowse function of the web-based tool OryGenesDB available on CIRAD 'South Green' Platform (orygenesdb.cirad.fr) which encompasses molecular resources (gene annotation, coding sequence or protein) of several grasses (rice, maize, sorghum). Putative gene function was then searched by identifying homologous genes from *Arabidopsis thaliana* genome using TAIR database (www.arabidopsis.org). Finally, we searched among these *A. thaliana* homologs for genes potentially regulated upon pathogen infection using PathoPlant website (www.pathoplant.de) that hosts microarray expression data involved in plant defence responses (Bülow et al. 2007).

Results

Analysis of SCYLV resistance traits in the cultivar panel

SCYLV resistance was assessed in the studied cultivar panel exposed to natural inoculation in the epidemiological context of Guadeloupe. Table 1 summarizes the estimates of the variance components for both ANOVA model 1 (two trials together) and ANOVA model 2 (separate trial analyses). Genetic variances (σ_G^2) and residual variances (σ_e^2) were always highly significant ($P < 0.0001$) for all resistance traits in both models. Their respective ratio (σ_G^2/σ_e^2) ranged between modest values for stalk resistance parameters (from 1.10 to 2.20) to higher values for leaf resistance parameters (from 2.53 to 4.40). In model 1, variances of genotype × trial interaction (σ_{GT}^2) and of genotype × crop cycles interactions (σ_{GC}^2) were highly significant ($P < 0.001$ or < 0.0001) for all traits. In the model 2, σ_{GC}^2 was highly

Fig. 1 Means of SCYLV resistance traits relative to leaf and stalk samplings in the cultivar panel, in plant cane (PC) and in second ratoon (R2) crop cycles in trials A and B for: virus incidence (**a**) and virus density (**b**)
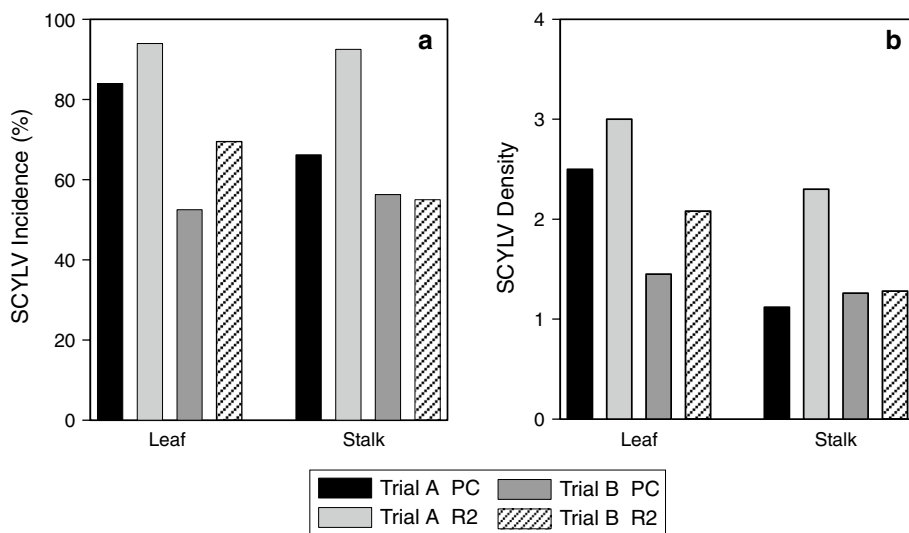

Table 2 Pearson correlation coefficients between resistant traits using cultivar estimated values

Trial	Trait	Plant part	A				B			
			SCYLV incidence		SCYLV density		SCYLV incidence		SCYLV density	
			Leaf	Stalk	Leaf	Stalk	Leaf	Stalk	Leaf	Stalk
A	SCYLV incidence	Leaf	1	0.83	0.95	0.72				
		Stalk		1	0.81	0.94				
	SCYLV density	Leaf			1	0.76				
		Stalk				1				
B	SCYLV incidence	Leaf	0.75	0.59	0.74	0.53	1	0.95	0.97	0.89
		Stalk	0.71	0.58	0.70	0.55		1	0.92	0.96
	SCYLV density	Leaf	0.71	0.57	0.75	0.55			1	0.90
		Stalk	0.67	0.61	0.68	0.63				1

All correlations are significant at $P < 0.0001$. Correlations within trials (A, B) are above the diagonal and correlations between trials are below the diagonal

significant in trial A for the four resistance traits (at least $P < 0.001$) but failed to be significant in two cases in trial B (stalk SCYLV incidence and density). Broad-sense heritability (H^2) at the level of both trials (model 1) estimated on an entry-mean basis using three replicates, two crop cycles and two trials ranged between 0.70 and 0.84. Due to the significant variance of interaction of genotype by trial, H^2 values computed at the experimental level of each trial (model 2) showed a range of 0.83–0.95 that is slightly higher than the former range. Despite small differences between models, all heritability values are high to very high. They indicate that both experimental design and phenotyping protocols provided accurate estimates of the genetic values of the cultivars surveyed and justify testing cultivar estimates produced by both ANOVA models for marker-trait association tests.

Cultivar panel means of disease resistance traits in each crop cycle studied for each trial are presented in Fig. 1. When comparing cycle within trials, the mean of foliar

disease traits consistently increased upon crop cycles in both trials (t test values ranging from -9.6 to -12.4 with $t = -9.6$; $P < 0.0001$). For stalk disease traits, the increase was significant in trial A (t values of -16.53 and -25.7 with $P < 0.0001$ for virus incidence and density, respectively) but not in trial B (t value of 0.62 and 0.817; $P = 0.53$ and 0.81 for virus incidence and density, respectively). Comparing crop cycle between trials, except for stalk density in plant cane, disease traits were significantly higher in trial A than in trial B for the same cycle with range of t value of 2.69 to 15.09 with respective P of 0.007–<0.0001.

Correlations between all traits (within and between trials) using cultivar PhV of model2 were highly significant ($P < 0.0001$). Pearson correlation coefficients varied from 0.72 to 0.97 for trait correlations within trials and from 0.53 to 0.75 for trait correlations between trials (Table 2). Relative good correlations between trials were found for SCYLV incidence ($r = 0.75$) and for SCYLV density

Table 3 Inflation factors computed for different genome-wide association models assessed on SCYLV resistance traits

Model co-factors	ANOVA		SCYLV incidence		SCYLV density	
	Model	Trial	Leaf	Stalk	Leaf	Stalk
Generalized linear model (GLM)						
Q2	1	A and B	1.211	1.118	1.194	1.051
Q1	1	A and B	1.848	1.708	1.761	1.466
Q2	2	A	1.134	1.088	1.106	1.071
Q1	2	A	1.827	1.557	1.735	1.334
Q2	2	B	1.117	1.103	1.202	1.074
Q1	2	B	1.710	1.541	1.524	1.395
Mixed linear model (MLM)						
Q2 + K	1	A and B	**0.976**	1.069	1.094	**1.011**
Q1 + K	1	A and B	1.153	1.117	1.173	**1.008**
Q2 + K	2	A	1.061	**1.009**	**1.002**	**1.007**
Q1 + K	2	A	**1.037**	**1.000**	1.061	**0.951**
Q2 + K	2	B	1.082	1.073	1.122	**1.020**
Q1 + K	2	B	1.153	1.170	1.223	1.146

Q1- and Q2-matrix are significant axes of PCA analyses inferred either from 619 independent DArT haplotypes/markers or from all 1,509 DArT markers, respectively. They represent a fixed co-factor (Q) for population structure. Co-factor K is the genetic similarity matrix. Inflation factors lower than 1.05 are indicated in bold characters

($r = 0.75$) in leaf tissues, whereas correlations between trials for SCYLV incidence ($r = 0.58$) and for SCYLV density ($r = 0.63$) in stalk tissues were lower. Correlations of traits within trials, between plant parts (leaf and stalk) were always high: 0.83 and 0.95 for SCYLV incidence in trials A and B, respectively, and 0.76 and 0.90 for SCYLV density in trials A and B, respectively.

Marker-trait associations

Structure analysis of the cultivar panel with PCA based on the independent 619 DArT haplotypes/markers yielded 11 axes that were significant ($P < 0.05$) with the Tracy–Widom test (T–W). Alternatively PCA with all 1,509 DArT markers yielded 24 significant other axes (T–W test; $P < 0.05$). Fitting each of the 12 estimates of resistance trait (SCYLV incidence and SCYLV density on leaves and stalks in two separate or combined trial analyses) with an increasing number of significant axes showed that minimum BIC was always achieved with only one axe in the first PCA and always three axes in the second PCA. Depending on resistance traits and ANOVA models considered, the first axis of the first PCA explained between 0.6 and 1.3 % of the phenotypic variation and the first three axes of the second PCA explained between 9.8 and 18.6 % of the phenotypic variation. Therefore, the two Q1 and Q2 alternative matrix used as covariate in GLM and MLM models always consisted of the first axis of the first PCA and the first three axes of the second PCA, respectively. Quantile–quantile plots of the 48 different genome-wide association exercises are presented in supplementary material 2.

Inflation factors λ of the all associations models are summarized in Table 3. Obviously, whatever the trait, in GLM models test statistics are largely more inflated when using Q1 than when using Q2, since λ ranged between 1.395 and 1.848 with Q1 compared to a lower range of 1.051–1.211 with Q2. In all GLM models, λ was always above the critical value of 1.05 defined by Price et al. (2010). Therefore, GLM models were not further considered in the presentation of marker-trait association results. As expected, MLM models showed a range of inflation factors (0.951–1.153) lower than in GLM models (1.051–1.848). Models Q2 + K gave lower λ values than Q1 + K in a majority of 8 cases over 12. All SCYLV resistance traits together, a total of 10 different MLM models exhibited an efficient control of inflation risk (λ < 1.05). Association results further discussed are only based on these 10 MLM models.

Table 4 presents the results of these ten GWAS exercises which correspond to different disease parameters, different trial analyses and MLM models. Considering all disease resistance traits and MLM models together, a total of six independent markers were involved in significant associations at a $P_{GWER} < 0.10$. These six independent markers corresponded to three AFLP (aggctc35, agccag14, agccta52) and three DArT markers (B424681, B424690, B424752). Among the six detected markers, only one (B424681) was located on the R570 reference genetic map (homology group VI). The three AFLP markers and B424752 DArT marker were significant in one analysis for a particular trait and MLM model. These four markers showed in one or two alternative trait analyses a −logP values higher than 4.00 but below the statistical threshold (4.28). B424681 and B424690 were significant in five and four different trait analyses, respectively. For leaf incidence with the Q1 + K MLM model, these two markers were significant for a −logP values of 6.09 and

Table 4 Results of marker-trait associations assessed on SCYLV resistance traits for MLM models showing an efficient control of inflation risk ($\lambda < 1.05$)

| Co-factors | ANOVA | | Disease parameter | | Markers (frequency) | | | | | |
| | Model | Trial | Trait | Plant part | $-\log P$ (R^2) | | | | | |
					Aggctc35 (0.06)	Agccag14 (0.17)	Agccta52 (0.10)	B424681 (0.20)	B424690 (0.08)	B424752 (0.14)
Q2 + K	2	A	Incidence	Stalk	4.05 (0.09)	2.66 (0.05)	3.50 (0.07)	**4.37 (0.09)**	**4.69 (0.10)**	2.84 (0.06)
Q2 + K	2	A	Density	Leaf	2.63 (0.05)	**4.71 (0.10)**	1.95 (0.03)	**4.84 (0.10)**	**4.33 (0.09)**	**5.25 (0.11)**
Q2 + K	2	A	Density	Stalk	3.88 (0.08)	2.66 (0.05)	3.48 (0.07)	3.39 (0.07)	4.01 (0.08)	2.84 (0.06)
Q2 + K	2	B	Density	Stalk	2.54 (0.05)	3.59 (0.07)	2.07 (0.04)	4.15 (0.09)	1.93 (0.03)	1.74 (0.03)
Q2 + K	2	A and B	Incidence	Leaf	3.58 (0.06)	4.01 (0.08)	1.65 (0.02)	2.86 (0.05)	1.58 (0.02)	4.19 (0.08)
Q2 + K	2	A and B	Density	Stalk	3.81 (0.08)	3.72 (0.08)	3.22 (0.06)	**4.47 (0.09)**	3.35 (0.06)	2.67 (0.05)
Q1 + K	2	A	Incidence	Leaf	3.40 (0.07)	3.87 (0.08)	4.00 (0.08)	**6.09 (0.14)**	**6.21 (0.14)**	3.99 (0.08)
Q1 + K	2	A	Incidence	Stalk	**4.47 (0.10)**	2.86 (0.06)	**4.29 (0.09)**	**4.85 (0.11)**	**4.65 (0.10)**	2.82 (0.06)
Q1 + K	2	A	Density	Stalk	3.97 (0.09)	3.01 (0.06)	4.17 (0.09)	3.59 (0.08)	3.78 (0.08)	3.05 (0.06)
Q1 + K	1	A and B	Density	Stalk	3.62 (0.08)	4.00 (0.09)	3.65 (0.08)	3.67 (0.08)	3.10 (0.06)	3.20 (0.06)

In bold character $-\log(P) > 4.28$ corresponding to an $P_{GWER} < 10\%$ (see material and method section). See legend of Table 3 for the signification of co-factors abbreviations

6.21 which corresponded to P_{GWER} values of 0.0015 and 0.0012, respectively.

Frequency of these six markers was relatively low, ranging from 0.06 to 0.20. All alleles showed a favorable effect on SCYLV resistance. According to the modest to high positive correlations between disease resistance traits between and within trials (Table 2), the associations were always in the same direction across all trait analyses, whatever their statistical significance, and even when the percentage of variation (R^2) explained was almost null.

The proportion of the total phenotypic variation (R^2) explained individually by significant markers in all trait MLM analyses ranged from 9 to 14 %.

Considering the distribution of the six detected resistant markers within the cultivar panel, number of markers present in individual cultivar ranged from zero to six markers. The number of cultivars bearing from zero to six markers reached 105, 44, 23, 11, 0 and 1 individuals, respectively. Due to their low frequency, cultivars bearing four to six markers were grouped into a single class for further statistic comparisons. As a rule, mean disease incidence values of the five cultivar classes bearing an increasing number of markers (0, 1, 2, 3 and 4–6 markers) decreased gradually: 88, 73, 52, 38 and 16 %, respectively, for leaf incidence; 79, 65, 51, 36 and 24 %, respectively for stalk incidence. A t test for each trait showed significant ($P < 0.5$) differences between all classes, except the fourth one that was not significantly different from its two neighbor classes. Moreover, the 35 cultivars bearing multiple markers (from two to six markers) were widely distributed among the 26 different geographical breeding origins represented in the cultivar panel.

The cumulative effects on disease resistance variations of the 6 detected markers were estimated with stepwise multiple regression models (SMRM) using the 176 cultivars with no missing information on detected marker (Table 5). The SMRMs captured between three and five of the six markers depending on trait and trials considered. The percentages of the variation of the panel explained by the markers retained in regression model varied from 20 to 31 %. The most frequently detected marker (B424681) in the GWAS analyses (Table 5) was kept for all disease resistance traits by the SMRM. Interestingly, three markers (aggctc35, agccag14 and B424752) detected only in one MLM trait model (Table 4) were retained in almost all regression models whatever the trait.

Search for resistance genes

The sequences of DArT markers B424681, B424690 and B424752 were blasted on the *Sorghum bicolor* genome sequence. Marker B424690 (526 bp) aligns on *S. bicolor* chromosome 8 with convincing BLAST parameters (HSP

Table 5 Results of stepwise multiple regressions of markers on SCYLV resistance traits in leaves and stalks, in both trials

Trial	Plant part	Phenotype	R^2	Aggctc35	Agccag14	Agccta52	B424681	B424690	B424752
A	Leaf	Incidence	0.31	X	X		X	X	X
A	Leaf	Density	0.31	X	X		X	X	X
A	Stalk	Incidence	0.26	X		X	X	X	X
A	Stalk	Density	0.25	X		X	X	X	X
B	Leaf	Incidence	0.22	X	X		X		X
B	Leaf	Density	0.21	X	X		X		X
B	Stalk	Incidence	0.22	X	X		X		X
B	Stalk	Density	0.20	X	X		X		X
A + B	Leaf	Incidence	0.28		X		X		X
A + B	Leaf	Density	0.29	X	X		X		X
A + B	Stalk	Incidence	0.27	X	X		X		X
A + B	Stalk	Density	0.26	X	X		X		X

R^2 phenotypic variation explained by the model (markers). Markers kept in the stepwise model are indicated by an X

bit score of 852 and an e-value of 0 with 95 % identity). The B424681 (569 bp) sequence showed 92 % sequence identity with a sequence located on *S. bicolor* chromosome 2 with a HSP bit score of 775 and an e-value of 0. The B424752 sequence of 540 bp, that contains a fragment of a Non-LTR retrotransposon LINE L1, could not be assigned to any specific sorghum chromosome due to partial alignment with multiple sorghum sequences. Figure 2 shows the sets of genes found in *Sorghum* in a region of 100 kb around the sorghum homologs of B424681 and B424690. Very close to B424681 sorghum homolog (at a distant of 11.8 kb), we found a gene homologous to the AT1G34030 gene of *A thaliana* that codes for the 40S ribosomal protein (RPS18B). At a further distance apart from B424681 sorghum homologs (19.9 kb and 78.7 kb), we also recognized two genes whose homologs have unknown functions in *A. thaliana* which are regulated upon the Tobacco mosaic virus (TMV) infection. Another TMV regulated homologous gene with unknown function was found at a distance of 40.7 kb from the B424690 sorghum homolog. Moreover, a gene homologous to a gene coding for a lipoxygenase (LOX5) related to aphid feeding in *A. Thaliana* (AT3G22400) was identified at only 6.4 kb from B424690 sorghum homolog.

Discussion

To identify genomic regions involved in the resistance of sugarcane to *SCYLV* disease, we studied associations within a panel of 189 modern sugarcane cultivars representative of the worldwide breeding germplasm. Phenotypic assessment of the panel for disease resistance was based on immunological diagnoses of the virus performed on material exposed to natural infections prevailing in the fields in Guadeloupe. To guarantee accurate phenotypic disease rating data in a context of variable natural epidemiological pressures, we invested in two successive repeated field trials phenotyped in two crop cycles using numerous sampling for SCYLV diagnosis. Due to the lack of the literature about the dynamic of the virus within infected sugarcane relative to its circulation and maintenance, we choose to diagnose both leaf and stalk tissues, to grasp a comprehensive picture of the susceptibility level of the cultivar panel and make sure of a proper biological interpretation of the resulting association study.

Disease resistance assessment relying on natural infection

In both trials, the level of SCYLV contamination was expected to be the result of background infection in cane seeds and further infections caused by SCYLV transmission to plant by the aphid vector. Means of disease trait values were significantly higher in the first trial (trial A) compared to the second trial (trial B) despite the same origin of cane seeds. This difference is attributed to the fact that aphid populations were profuse only in the first crop cycle of trial A, but sparsely observed the other times. Nevertheless, broad-sense heritability values at trial level of all disease resistance traits appeared always high to very high (0.80–0.95) in both trials and also in their combined analysis (0.70–0.84). These findings show efficient control of environmental sources of variations within trial when estimating the genetic values of the cultivars surveyed. A high average value in virus incidence (75 %) was observed in leaves across trials and crop cycles. This level of incidence is similar to incidence values observed in cultivars surveyed in Hawaii (Komor et al. 2010) and Florida (Comstock et al. 1999) that had been exposed for many years to potential virus infection.

Association tests for SCYLV resistance

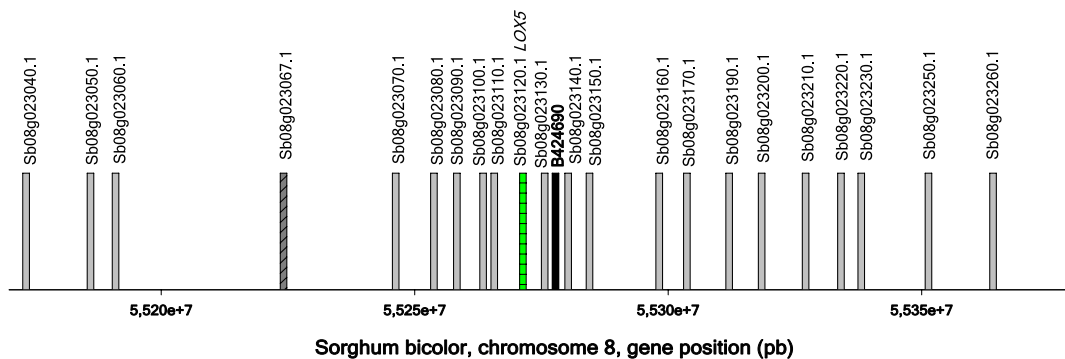Phenotype data from each trial were used to tag for SCYLV resistance alleles through GWAS approaches. In GWAS,
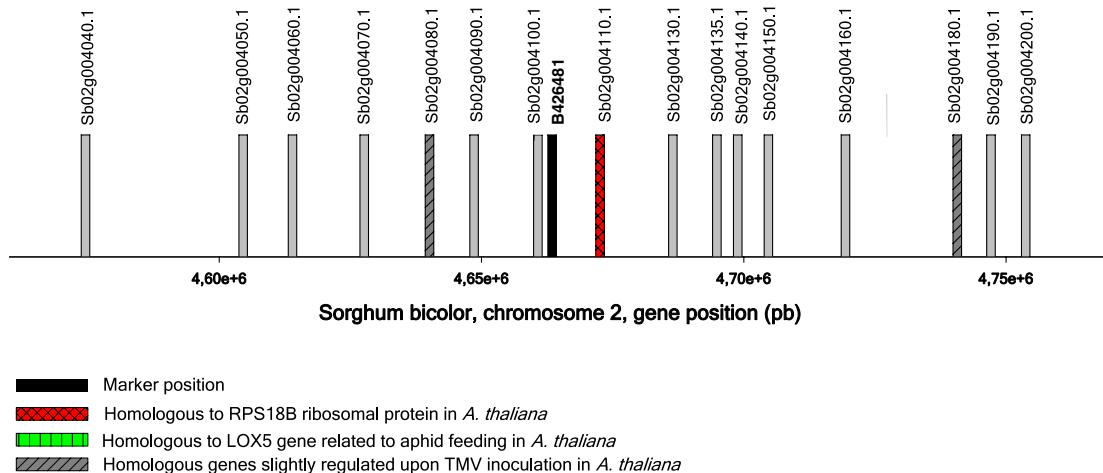
Fig. 2 Genes in the sorghum regions ($d = 100$ kb) homologous to two of the sugarcane regions found associated with SCYLV resistance

population structure has to be correctly modeled to control efficiently the risk of spurious associations (Patterson et al. 2006; Yu et al. 2006). In the polyploid context of sugarcane, the use of dominant markers whose geneticists still depend on, such as the ones we used, constitute an obvious handicap to model accurately population structure. In our study, we attempted a general linear model (GLM) with two alternative population structure co-factors derived from a principle component analysis (PCA) using either all DArT polymorphic markers available or a sampling strategy of independent ones. Both GLM association exercises failed to control an inflation risk of test statistics and possible spurious associations. On the contrary, mixed linear models (MLM) with the addition of a kinship matrix covariate, allowed a strict control of the inflation of test statistic ($\lambda < 1.05$) on ten occasions.

These ten MLM genome-wide association studies revealed six independent significant markers linked to one or several SCYLV resistance parameters at a genome-wide error rate (GWER) lower than 10 %. Among them

two DArT markers (B424681 and B424690) were detected at a P_{GWER} of 0.0015 and 0.0012 for a particular resistance trait (disease leaf incidence). All the six markers were associated with resistance. These six markers were present in the cultivar panel at a low frequency (6–20 %). The low frequency of resistance alleles in our worldwide cultivar panel combined with the high virus incidence mean reflects (1) the absence of selection in breeding programs due to the recent spread of the disease (Komor et al. 2010) and (2) a probable relative scarcity of sources of resistance available in the modern interspecific sugarcane germplasm. Individual allele effects associated with yellow leaf resistance traits explained from 9 to 14 % of the total variation of the cultivar panel. These effect sizes which are overestimated regarding the modest panel size of the panel were similar to those obtained in previous sugarcane association study of Wei et al. (2006) relative to three sugarcane disease resistance (Fiji disease, leaf-scald and pachymetra root rot) and based on a small panel (154 accessions).

Multiple regression analysis showed that the part of variation in yellow leaf resistance traits, measured by leaf diagnosis, varied from 20 to 31 % depending on trait assessed. Our results are similar to Wei et al. (2006) association study which detected in sugarcane four markers relative to the resistance of another vectored virus disease (Fiji disease) explaining together 32 % of the variation in the cultivar panel surveyed. The large part of the variation yet left unexplained might be attributed to the addition of several reasons: (1) a certain number of perceptible alleles of interest might have been missed due to an insufficient marker density; (2) QTL of medium effect in a single variety might exhibit a too small quantitative effect in the context of the association study to be detected whatever marker density; (3) rare alleles of significant effects in a few cultivar might have also been missed and fell in the "missing heritability" category (Yang et al. 2010); (4) finally, depending on the narrow-sense heritability of the disease traits, an undetermined and inaccessible proportion of the variance might be due to more or less complex allele interactions.

The modest part of variation explained in our study could be explained by a lack of power in our association tests. Four main factors limited the detection of markers: (1) the modest size of the panel: it allowed tagging a few alleles that are believed to be some of the most important ones. A larger panel should substantially improve the statistical power of detection of medium- and small-marker effects (Long and Langley 1999). (2) The confounding effect of polyploidy: in polyploid crop species, detection of linkage disequilibrium (LD) between a QTL and a marker in its vicinity could be missed due to unbalanced frequencies between QTL and marker. The risk of undetected genuine linkage is expected to increase with polyploidy level. In the context of the high ploidy level of sugarcane, such undetected LD was shown to be common with dominant markers (Fig. 3 in Raboin et al. 2008). (3) The use of dominant markers: their dichotomous nature (presence versus absence) involves a loss of power in QTL detection compared to co-dominant markers which enable homozygotes and heterozygotes to be distinguished. In out-crossing polyploidy species such as sugarcane, only markers segregating in very low dosage (simplex or duplex) behind the dominant 'presence' phenotypes are believed to be sufficiently informative to tag alleles of agronomic interest. The fact that the six markers found in our study were all in low frequency in the panel (0.06–0.20) gave credit to this intuitive believe. (4) Finally, the number of markers we used, despite substantial, was probably far from sufficient to tag all the haplotype diversity that segregate in our panel. Capturing the polymorphic fraction of the genome is especially challenging in the polyploid species, where copies of homologous chromosomes "dilute" the polymorphism.

These well-known constraints in the genetic analysis of polyploids that severely limit the power of QTL detection could be overcome with the advent of single-nucleotide polymorphism (SNP) technology and statistical innovations in data interpretation dedicated to polyploids. Recently, Hackett et al. (2013) adapted current methodologies of linkage and QTL analyses of diploid to autotetraploid potato using SNP dosage. Interestingly, Serang et al. (2012) developed a formal statistical method allowing SNP genotype calling and allele dosage estimates in polyploids. A Bayesian model based on comparative allele signal intensities, gives access to genotype configuration with any bi-allelic SNP loci of any sort of polyploids even of the complex sugarcane (Garcia et al. 2013). Moving from dominant markers to bi-allelic SNP with dosage information should improve the power of association studies due to: (1) the inference of more realistic genetic structure likely reducing the inflation of test statistics, (2) the opportunity to conceive refined genetic model to analyze within loci effects.

Candidate genes in the vicinity of DArT markers

Resistance of plant to a vectored virus might result from antixenosis or antibiosis phenomenon as observed in sugarcane against *M. sacchari* (Fartek et al. 2012) or from the limitation of the multiplication and the movement of the virus within the host (Kang et al. 2005). We had the opportunity to search for genes in the vicinity of two significant DArT markers that blasted on a unique locus in the genome of sorghum. In the sorghum regions homolog to the sugarcane regions associated with SCYLV resistance, we discover a few genes involved in Arabidopsis aphid or virus interaction.

One of these genes encodes for a particular lipoxygenase (LOX5) favoring aphid feeding and colonization of Arabidopsis foliage by a peach aphid species (Nalam et al. 2012). Lipoxygenase (LOXs) multi gene family catalyze hydroperoxidation of free fatty acids into diverse biologically actives compounds named oxylipins. Interestingly, some isoforms of LOX genes are known to be involved in defence response to pathogens by producing signaling compounds and antimicrobial compounds (Liavonchanka and Feussner 2006). Moreover, a LOX isoform recently characterized in tea plant was shown to be involved in plant response following tea feeding by phloem-feeder pests among which an aphis species (Liu and Han 2010).

Another candidate gene that we detected is a homolog of *A. thaliana* gene encoding a 40S ribosomal protein S18 (RPS18B). Interestingly the 40S ribosomal subunit is involved in the initiation of the translation of viral RNA. Virus multiplication depends on the translation machinery of its host and interferences with numerous specific mechanism generated by the virus that could be part of

the host-range control of plant viruses (Fütterer and Hohn 1996; Thiebeauld et al. 2007). The hypothesis that this corresponding gene is implicated in virus population regulation is congruent with the high frequency of the detection of the corresponding marker (B424681) by GWAS and multiple regression analysis.

Finally, among the candidate genes detected we found three genes, homologous to genes with unknown functions that are regulated upon the Tobacco mosaic virus (TMV) infection in *A. thaliana*.

Prospects about breeding applications

Previously, the unique genetic study carried out to characterize yellow leaf resistance in sugarcane was on the basis of a biparental cross between a susceptible and a resistant cultivar (Costet et al. 2012). A major quantitative resistance allele (QRA) to SCYLV, named *Ryl*1, was tagged in the resistant parent (MQ76-53) thanks to three AFLP markers. This resistant clone represents an atypical semi-exotic sugarcane genotype since it is derived from a cross between the old domesticated sugar-producing line 'Trojan' (CO 270 × *S. officinarum*) and the wild accession (*S. spontaneum*) SES 528 (Raboin et al. 2006). MQ76-53 was present in our association panel, but none of the six resistant markers we detected was detected in MQ76-53 genome. As mentioned by Costet et al. (2012), the most significant markers associated with *Ryl*1 in MQ76-53, were absent in the small cultivar panel ($n = 72$) studied by Raboin et al. (2008), that was representative of the worldwide current sugar-producing breeding germplasm. There hence is a high chance that *Ryl*1 is present at a very low frequency in our current panel that is also representative of the worldwide sugar-producing germplasm used in breeding, and, the probability of the detection of this major resistance gene is believed to be unlikely.

Among the six markers detected, two markers (B424681 and B424690) were repeatedly found significant ($P_{GWER} < 0.01$) up to highly significant ($P_{GWER} < 0.0015$). Regarding their effect size, these loci have a high chance to represent major resistance alleles in most of the cultivars that bear them. Some of markers found in this study might also tag genuine alleles of resistance with significant quantitative effect in a single cultivar genome. It would be interesting to test the significance of these candidate markers in ad hoc QTL studies between polymorphic parents for each marker to check the existence of related major alleles and evaluate their effect size. Identifying new accessions bearing additional major resistant QTLs different from *Ryl*1 in germplasm less exotic than MQ76-53, would allow for greater flexibility in breeding programs as well as allowing sustainable pyramid strategies with different resistance sources.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Bülow L, Schindler M, Hehl R (2007) PathoPlant®: a platform for microarray expression data to analyze co-regulated genes involved in plant defense responses. Nucleic Acids Res 35:841–845

Comstock JC, Irey MS, Lockhart BEL, Wang ZK (1998) Incidence of yellow leaf syndrome in CP cultivars based on polymerase chain reaction and serological techniques. Sugar Cane 4:21–24

Comstock JC, Miller JD, Tai PYP, Follis JE (1999) Incidence of and resistance to sugarcane yellow leaf virus in Florida. Proc Int Soc of Sugar Cane Technol 23:366–372

Comstock JC, Miller JD, Schnell RJ (2001) Incidence of sugarcane yellow leaf virus in clones maintained in the world collection of sugarcane and related grasses at the United States national repository in Miami, Florida. Sugar Tech 3(4):128–133

Costet L, Raboin LM, Payet M, D'Hont A, Nibouche S (2012) A major quantitative trait allele for resistance to the Sugarcane yellow leaf virus (Luteoviridae). Plant Breed 131:637–640

D'Arcy CJ, Domier LL (2005) Luteoviridae. In: Fauquet MC, Mayo MA, Maniloff J, Desselberger U, Ball LA (eds) Virus Taxonomy. VIIIth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, New York, pp 891–900

D'Hont A, Ison D, Alix K, Roux C, Glaszmann JC (1998) Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of RNA genes. Genome 41:221–225

Daugrois JH, Jean-Baptiste I, Lockhart BEL, Irey S, Chatenet M, Rott P (1999) First report of sugarcane yellow leaf virus in French West Indies. Plant Dis 83:588

Daugrois JH, Edon-Jock C, Bonoto S, Vaillant J, Rott P (2011) Spread of Sugarcane yellow leaf virus in initially disease-free sugarcane is linked to rainfall and host resistance in the humid tropical environment of Guadeloupe. Eur J Plant Pathol 129:71–80

Davis MJ, Dean JL, Miller JD, Shine JM (1994) A method to screen for resistance to ratoon stunting disease of sugarcane. Sugarcane 6:9–16

Edon-Jock C, Rott P, Vaillant J, Fernandez E, Girard JC, Daugrois JH (2007) Status of sugarcane yellow leaf virus in commercial fields and risk assessment in Guadeloupe. Proceedings of the XXVI congress of International Society of Sugarcane Technologists, Durban, South Africa, 2007. Proc Int Soc Sugar Cane Technol. 26:995–1004

Fartek B, Nibouche S, Turpin P, Costet L, Reynaud B (2012) Resistance to Melanaphis *sacchari* in the sugarcane cultivar R 365. Entomol Exp Appl 144(3):270–278

Fütterer J, Hohn T (1996) Translation in plants—rules and exceptions. Plant Mol Biol 32:159–189

Garcia AAF, Mollinari M, Marconi TG, Serang O, Silva RR, Vieira MLC, Vicentini R, Costa EA, Mancini M, Garcia MO, Pastina MM et al (2013) SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. Sci Rep 3:3399

Garsmeur O, Charron C, Bocs S, Jouffe V, Samain S, Couloux A, Droc G, Zini C, Glaszmann JC, Van Sluys MA, D'Hont A (2010) High homologous gene conservation despite extreme autopolyploid redundancy in sugarcane. New Phytol 168:629–642

Hackett CA, Mclean K, Bryan GJ (2013) Linkage analysis and QTL mapping using SNP dosage data in a tetraploid potato mapping population. PLoS One 8:e6339

Hall DG, Bennett FD (1994) Biological control and IPM of sugarcane pests in Florida. In: Rosen D, Bennett FD, Capinera JL (eds.), Pest management in the subtropics, biological control—a Florida perspective. Intercept Ltd., Andover, United Kingdom, pp. 287–325

Heller-Uszynska K, Uszynski G, Huttner E, Evers M, Carlig J, Caig V, Aitken K, Jackson P, Piperidis G, Cox M, Gilmour R, D'Hont A, Butterfield M, Glaszmann JC, Kilian A (2011) Diversity arrays technology effectively reveals DNA polymorphism in a large and complex genome of sugarcane. Mol Breed 28:37–55

Hoarau JY, Offmann B, D'Hont A, Risterucci AM, Roques D, Glaszmann JC, Grivet L (2001) Genetic dissection of a modern sugarcane cultivar (Saccharum spp.). I. Genome mapping with AFLP markers. Theor Appl Genet 103:84–97

Jannoo N, Grivet L, Dookun A, D'Hont A, Glaszmann JC (1999) Linkage disequilibrium among modern sugarcane cultivars. Theor Appl Genet 99:1053–1060

Jannoo N, Grivet L, Chantret N, Garsmeur O, Glaszmann JC, Arruda P, D'Hont A (2007) Orthologous comparison in gene-rich region among grasses reveals stability in the sugarcane polyploid genome. Plant J 50:574–585

Kang BC, Yeam I, Jahn MM (2005) Genetics of plant virus resistance. Annu Rev Phytopathol 43:581–621

Komor E (2011) Susceptibility of sugarcane, plantation weeds and grain cereals to infection by Sugarcane yellow leaf virus and selection by sugarcane breeding in Hawaii. Eur J Plant Pathol 129:379–388

Komor E, ElSayed A, Lehrer AT (2010) Sugarcane yellow leaf virus introduction and spread in Hawaiian sugarcane industry: retrospective epidemiological study of an unnoticed, mostly asymptomatic plant disease. Eur J Plant Pathol 127:207–217

Liavonchanka A, Feussner I (2006) Lipoxygenases: occurrence, functions and catalysis. J Plant Physiol 163:348–357

Liu S, Han B (2010) Differential expression pattern of an acidic 9/13-lipoxygenase in flower opening and senescence and in leaf response to phloem feeders in the tea plant. BMC Plant Biol 10:228

Lockhart BEL, Cronjé CPR (2000) Yellow leaf syndrome. In: Rott P, Bailey RA, Comstock JC, Croft BJ, Saumtally AS (eds) A guide to sugarcane diseases. La Librairie du CIRAD, Montpellier, pp 291–295

Long AD, Langley CH (1999) The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. Genome Res 9:720–731

Nalam VJ, Keeretaweep J, Sarowar S, Shah J (2012) Root-derived oxylipins promote green peach aphid performance on *Arabidopsis* foliage. Plant Cell 24:1643–1653

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2:2074–2092

Perrier X, Jacquemoud-Collet JP (2006) DARwin software. http://darwin.cirad.fr/darwin. Accessed 04 Feb 2014

Price AL, Zaitlen NA, Reich D, Patterson N (2010) New approaches to population stratification in genome-wide association studies. Nat Rev Genet 11(7):459–463

Raboin LM, Oliveira KM, Le Cunff L, Telismart H, Roques D et al (2006) Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: identification of a gene controlling stalk colour and a new rust resistance gene. Theor Appl Genet 112:1382–1391

Raboin LM, Pauquet J, Butterfield M, D'Hont A, Glaszmann JC (2008) Analysis of genome-wide linkage disequilibrium in the highly polyploidy sugarcane. Theor Appl Genet 116:701–714

Rassaby L, Girard JC, Letourmy P, Chaume J, Irey MS, Lockhart BEL, Kodja H, Rott P (2003) Impact of Sugarcane yellow leaf virus on sugarcane yield and juice quality in Réunion Island. Eur J Plant Pathol 109:459–466

Rott P, Mirkov TE, Schenck S, Girard JC (2008) Recent advances in research on Sugarcane yellow leaf virus, the causal agent of sugarcane yellow leaf. Sugar Cane Int 26:18–27

Schenck S (2001) Sugarcane yellow leaf syndrome: history and current concepts. In: Rao GP, Ford RE, Tosic M, Teakle DS (eds) Sugarcane pathology, vol II., Virus and phytoplasma diseasesScience Publisher, Enfield, pp 25–35

Schenck S, Lehrer AT (2000) Factors affecting the transmission of Sugarcane yellow leaf virus. Plant Dis 84:1085–1088

Schenck S, Hu JS, Lockhart BEL (1997) Use of a tissue blot immunoassay to determine the distribution of Sugarcane yellow leaf virus in Hawaii. Sugar cane 4:5–8

Serang O, Mollinari M, Garcia AAF (2012) Efficient exact maximum a posteriori computation for Bayesian SNP genotyping in polyploids. PLoS One 7:e30906

Sreenivasan TV, Ahloowalia BS, Heinz DJ (1987) Cytogenetics. In: Heinz DJ (ed) Sugarcane improvement through breeding. Elsevier, Amsterdam, pp 143–210

Thiebeauld O, Pooggin MM, Ryabova LA (2007) Alternative translation strategies in plant viruses. Plant Viruses 1:1–20

Vega J, Scagliusi SMM, Ulian EC (1997) Sugarcane yellow leaf disease in Brazil: evidence of association with a luteovirus. Plant Dis 81:21–26

Wang J, Roe B, Macmil S, Yu Q, Murray J, Tang H, Chen C, Najar F, Wiley G, Bowers J, Van Sluys MA, Rokhsar D, Hudson M, Moose S, Paterson A, Ming R (2010) Microcollinearity between autopolyploid sugarcane and diploid sorghum genomes. BMC Genomics 11:261

Wei X, Jackson P, McIntyre C, Kilian A, Croft B (2006) Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. Theor Appl Genet 114:155–164

Wei X, Jackson PA, Hermann S, Kilian A, Heller-Uszynska K, Deomano E (2010) Simultaneously accounting for population structure, genotype by environment interaction, and spatial variation in marker-trait associations in sugarcane. Genome 53:973–981

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42:565–569

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu K, Arnett JD, Ordovas JM, Buckler ES (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–360

Zhou GH, Li JG, Xu DL, Shen WK, Deng HH (2006) Occurrence of Sugarcane yellow leaf virus in South China and its transmission by the sugarcane colonizing aphid, *Ceratovacuna lanigera*. Scientia Agricultura Sinica 39:2023–2027

Zhu C, Yu J (2009) Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. Genetics 182:875–888

Zhu YJ, Lim STS, Schenck S, Arcinas A, Komor E (2011) RT-PCR and quantitative real-time RT-PCR detection of *Sugarcane Yellow Leaf Virus* (SCYLV) in symptomatic and asymptomatic plants of Hawaiian sugarcane cultivars and the correlation of SCYLV titre to yield. Eur J Plant Pathol 127:263–273