American Journal of

Biochemistry and Molecular Biology

ISSN 2150-4210

Academic
Journals Inc.

www.academicjournals.com

OPEN ACCESS American Journal of Biochemistry and Molecular Biology

ISSN 2150-4210
DOI: 10.3923/ajbmb.2017.41.52

Research Article
Characterization of ORF0 and ORF1 and their Roles in Recombination and Replication of *Sugarcane yellow leaf virus*

[1]A.I. ElSayed, [2]M. Boulila, [3]D.C. Odero, [1,4]A.A. Omar, [1]A.M. Gomaa and [5]R. Viswanathan

[1]Department of Biochemistry, Faculty of Agriculture, Zagazig University, 44511 Zagazig, Egypt
[2]Institut de l'Olivier, B.P. 14, 4061 Sousse Ibn-khaldoun, Tunisia
[3]Everglades Research and Education Center, IFAS, University of Florida, 3200 East Palm Beach Road, Belle Glade, FL 33430, United State of America
[4]Citrus Research and Education Center, IFAS, University of Florida, 700 Experiment Station Road, Lake Alfred, FL 33850, United State of America
[5]Division of Crop Protection, Sugarcane Breeding Institute, Indian Council of Agricultural Research, 641007 Coimbatore, India

Abstract

Background: Two open reading frames (ORF0 and ORF1) of *Sugarcane yellow leaf virus* (SCYLV) genome which play an important role in virus replication and accumulation in plants have been characterized. **Materials and Methods:** The RT-PCR and qRT-PCR have been used to evaluate the degree of infection with SCYLV using different genome locations (ORF0 and ORF1). Possible recombination events in two ORFs were identified using TOPALi (v2.5), RECCO and RDP software's. **Results:** Transcript levels of two genes varied among infected plants but overall expression of ORF1 was higher than ORF0. Cultivar H73-6110 (susceptible) yielded the highest transcript levels of ORF1 whereas cultivar H78-4153 (resistant) exhibited the lowest levels. No significant differences were found between the sugarcane cultivars for the ORF0 transcripts in mature leaves and seedling tissues. Amino acid sequence similarity of ORF0 and ORF1 varied among SCYLV isolates and ranged from 69-99 and 73-99%, respectively. Possible recombination events located in the two ORFs were identified using TOPALi (v2.5), RECCO and RDP software's. **Conclusion:** The results showed strong presence of recombination in aligned sequences of ORF0 and ORF1 when TOPALi and RECCO programs were used.

Key words: Open reading frame, phylogenetic analysis, recombination, *Sugarcane yellow leaf virus*, qRT-PCR

Received: September 02, 2016 Accepted: October 13, 2016 Published: December 15, 2016

Citation: A.I. ElSayed, M. Boulila, D.C. Odero, A.A. Omar, A.M. Gomaa and R. Viswanathan, 2017. Characterization of ORF0 and ORF1 and their roles in recombination and replication of *Sugarcane yellow leaf virus*. Am. J. Biochem. Mol. Biol., 7: 41-52.

Corresponding Author: A.A. Omar, Department of Biochemistry, Faculty of Agriculture, Zagazig University, 44511 Zagazig, Egypt

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Sugarcane yellow leaf virus (SCYLV), the causative agent of yellow leaf disease (YLD) of sugarcane (*Saccharum* spp. hybrid) is a single stranded, positive sense, RNA genome, with six open reading frames (ORFs 0-5) and three untranslated regions (UTRs) consisting of ~5.8 kb nucleotides[1,2]. The three 5'-proximal ORFs are translated directly from the genomic RNA and include ORF1 which overlaps ORF0 and ORF2 in the 5' and 3' termini, respectively. The ORF0 encodes the SCYLV P0 protein, an RNA silencing suppressor[3]. The ORF1 and ORF2 are translated together, encoding a serine protease and a RNA dependent RNA polymerase (RdRp), respectively[1]. The ORF3 encodes a viral coat protein and ORF4 harbored inside ORF3 encodes a viral movement protein. In addition, ORF5 encodes a 52.1 kDa protein that probably is involved in aphid transmission of SCYLV[2,4].

The assumed amino acid sequence of P0 proteins is highly conserved among different geographic isolates of SCYLV[3,5]. In contrast, the amino acid sequence similarity between P0s of *Poleroviruses* is very low. The potato leaf roll virus (PLRV-P0) which has 21% sequence identity is the best match of SCYLV-P0 among any other *Polerovirus*-P0[3]. Generally, the role of ORF0 of *Luteoviruses* in the viral infection cycle has not been determined even though the translation product (p28-32) of this ORF has been suggested to play a role in host recognition[6]. In addition, ORF0 in potatoes was found to be important in PLRV symptom development[7].

The 5' region (ORF1 and ORF2) of the SCYLV genome might have originated from a *Polerovirus*, whereas the 3' region (ORF3 and ORF4) was derived from a *Luteovirus*. The region from ORF5 to the read through protein gene came from an *Enamovirus*[8]. This shows that SCYLV is a developing virus that evolved by the recombination of ancestors in the genera *Luteovirus*, *Polerovirus* and *Enamovirus* of the family Luteoviridae[9-11]. Thus, recombination is the dominant feature of *Poleroviruses* as well as of SCYLV. Although, several studies have reported sequencing and characterization of SCYLV genome[1,2,5,9-15], there is limited information on recombination of the SCYLV gene and its role in replication of the virus. Therefore, the purpose of the present study was to enrich our understanding regarding to recombination and selection pressure of SCYLV. Analyses accomplished herein were based on the sequences of the ORF0 and ORF1 coding genes.

MATERIALS AND METHODS

Plant material: Five sugarcane (*Saccharum* spp. hybrids) cultivars infested with SCYLV were grown in pots in a greenhouse and used for RNA isolation and purification. Sugarcane cultivars H73-6110 and H87-4094 (susceptible), H87-4319 and H78-4153 (resistant) and H65-7052 (intermediately susceptible)[16] originated from Hawaii and were provided by the Hawaiian Agriculture Research Center (Aiea, Hawaii, USA). The RNA was extracted using mature leaves and seedling of five sugarcane cultivars to perform the required analyses.

RT-PCR to evaluate the degree of infection with SCYLV using different genome locations (ORF0 and ORF1): Total RNA from mature leaves and seedling was isolated by SDS-phenol/chloroform extraction as described by ElSayed *et al.*[5]. Isolated RNA was used as template for RT-PCR in a PCR reaction using T100 Thermal Cycler (Bio-Rad, France). The specific primers were used to amplify ORF0 and ORF1 of the SCYLV genome for the five sugarcane cultivars are presented in Table 1. The RNA was reverse transcribed using the RevertAid H Minus First Strand cDNA Synthesis Kit (Fermentas GmbH, Leon-Rot, Germany) according to the manufacturer's protocol (priming with 50 pmol of reverse primers and 1 µg of RNA template). The PCR reaction was carried out in 25 µL containing 1 µL cDNA, 2.5 µL of 10x PCR buffer containing 15 mM $MgCl_2$, 0.5 µL of 10 mM dNTP mix, 0.1 µL (100 µM) of each forward and reverse primer, 1 unit of Taq polymerase (Stratagene, Waldbronn, Germany) and sterile milliQ water added up to the final volume of 25 µL. The PCR program comprised of initial denaturation step at 94°C for 4 min, 10 cycles at 94°C for 30 sec, 62°C for 2 min, 72°C for 45 sec, followed by 30 cycles at 94°C for 30 sec, 62°C for 30 sec and 72°C for 45 sec with a final extension at 72°C for 7 min. Optimization of the PCR program, including design of primer sets, was achieved by using semi qPCR with different numbers of PCR-cycles. The RT-PCR was accomplished with the 25S rRNA internal control that was used as a reference gene to normalize gene expression level and to estimate the

Table 1: Primers for RT-PCR and quantitative RT-PCR

Gene name	5'-3' primer sequence	T_A
qRT-PCR primers		
GAPDH	GTGGTGCCAAGAAGGATGTT	58°C
	GTTGTGCAGCTAGCATTGGA	
ORF0	ATGGTGCCTATTCTGCTCCT	58°C
	GCTTGGAACGGCATCTCTTA	
ORF1	AGCTCGTCATTGATCGTGTG	58°C
	CAGGAATTTGGGGTCTTCAA	
RT-PCR primers		
ORF0	TTTGGACCAAGCCTCTGACT	62°C
	GGCAAGCCATAAAAGGACAG	
ORF1	CAGACATTGCTGATTAC	62°C
	GCTCTCCACAAAGCTATCT	
25S rRNA	CGTGGCCTATCGATCCTTTA	62°C
	AACCTGTCTCACGACGGTCT	

integrity of the cDNA. A 10 µL aliquot of each amplified product was examined by electrophoresis on 1% agarose gels pre-stained with ethidium bromide (5 µL/100 mL of agarose solution).

Real time RT-PCR assay: Quantitative real-time PCR (qRT-PCR) analysis was carried out on the iCycler™ Thermal Cycler (Bio-Rad, USA) with the iQ™SYBR Green Supermix (Bio-Rad, USA) in a final volume of 20 µL according to the manufacturer's instructions. The iCycler was programmed to 95°C for 1 min; 45 cycles at 95°C for 30 sec, 58°C for 40 sec and 72°C for 45 sec and then final extension at 72°C for 10 min followed by a melting curve programme (55-95°C in increasing steps of 0.5°C). The GAPDH gene was used as control for the normalization of the qRT-PCR. Efficiencies of each reaction were calculated using LinReg PCR software[17]. Signal values were consequently derived from the threshold cycles (average background subtracted) using the equation of Pfaffl[18].

Alignment of sequences and construction of phylogenetic trees: A total of 72 nucleotide and amino acid sequences (available so far in the databases) of SCYLV (45 of ORF0 and 27 of ORF1) repossessed from GenBank were aligned using CLUSTALW 2.1, CLUSTALX 2.1[19] and Multalin[20] software programs with default settings. The phylogenetic relationships among SCYLV isolates were resolute with the Maximum Likelihood (ML) algorithm combined in the MEGA version 6 program[21]. Based on the evaluation of best fit substitution model implemented in MEGA6, the ML tree was reconstructed under the hypothesis of substitution model K2 coupled to a discrete Gamma distribution (+G) with five rate categories[22] for both ORF0 and ORF1 sequences. The substitution model parameters estimated for ORF0 were (i) Base frequencies: f (A) = f (T) = f (C) = f (G) = 0.250, (ii) Substitution rates: r (AT) = r (AC) = r (CG) = r (GT) = 0.033; r (AG) = r (CT) = 0.184 and (iii) Transition/transversion ratios: R = 2.93. The bayesian information criterion value (BIC = 7310.965) with K2+G (+G = 0.33) model was the lowest among the 24 models tested. The substitution model for ORF1 was also K2+G (+G = 0.23). The parameters projected for ORF1 were (i) Base frequencies: f (A) = f (T) = f (C) = f (G) = 0.250, (ii) Substitution rates: r (AT) = r (AC) = r (CG) = r (GT) = 0.029; r (AG) = r (CT) = 0.192 and (iii) Transition/transversion ratios: R = 3.31. The bayesian information criterion value (BIC = 13611.302) with K2+G model was the lowest among the 24 models tested. Bootstrap analyses were achieved with 1,000 replicates to evaluate the robustness of the tree branches.

Recombination analyses: Potential recombination events between diverged nucleotide sequences were explored with three different programs: RDP4.3[23,24], RECCO[25] and TOPALi v2.5 applying difference of sums of squares (DSS) statistics[26]. The RDP includes several published recombination detection methods into a single suite of tools: The RDP[27], GENECONV[28], BOOTSCAN[27], MAXCHI[29], CHIMAERA[30], SISCAN[31] and 3SEQ[32]. Some methods are phylogeny-based (BOOTSCAN, RDP and SISCAN) while the others are substitution-based methods (GENCONV, MAXCHI, CHIMAERA and 3SEQ). In all the cases, default parameters were used. Only events predicted by more than half of the methods were considered as significant. The algorithm established and described by Maydt and Lengauer[25] is simple and sensitive method for identifying recombination in a set of sequences and locating putative recombination breakpoints is based on cost minimization. This method has only two tunable parameters, recombination and mutation cost. In practice the only parameter considered is α representing the cost of mutation relative to recombination. When α changes from 0-1, the cost of mutation weighted by α increases and the cost for recombination weighted by 1-α decreases. Therefore, the parameter α controls the ambiguity between mutation and recombination. The TOPALi v2.5 incorporates the difference of sums of squares (DSS) statistics[26]. This method uses a sliding window and considers changes in the branching patterns of the trees estimated on the windows along the alignment, corresponding to high values of DSS.

RESULTS

Expression and transcript levels of ORF0 and ORF1 using RT-PCR and qRT-PCR: The expression of ORF0 and ORF1 was identified by RT-PCR using specific primers (Table 1) in infected mature leaves and seedling for the five sugarcane cultivars. Results in Fig. 1a and b are showing that the expression of ORF0 in mature leaves and seedling of the Hawaiian sugarcane cultivars were constant even in the cultivars that fluctuated in their infection rates[33] such as H65-7052 and H78-4153. However, these cultivars had relatively low ORF0 expression compared to the other cultivars.

Amplification of ORF1 also produced a band of the expected size (1200 bp) for all infected sugarcane cultivars (Fig. 2a, b), but difference in band intensity was clearly observed in both mature leaves and seedling. However, the band with the lowest intensity was observed in cultivar H78-4153 in the mature leaf. In the seedling, cultivars H65-7052 and H78-4153 yielded the lowest intensity bands (Fig. 2b).

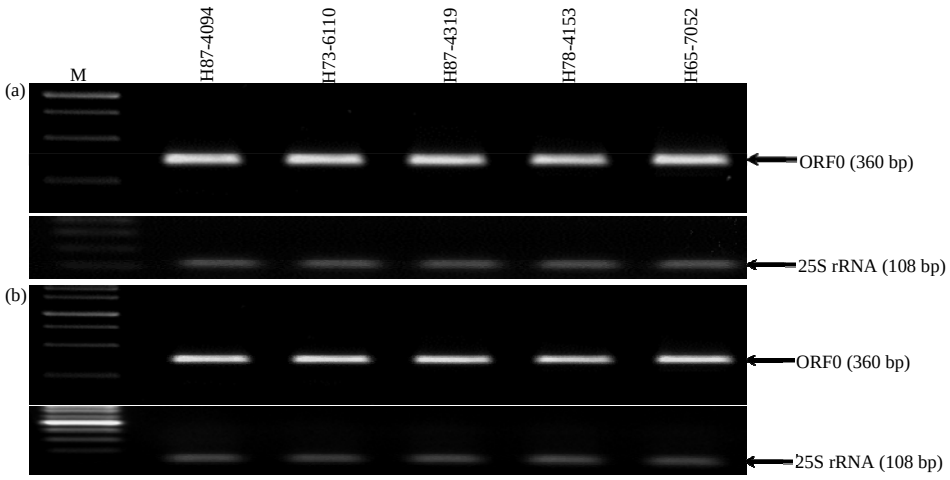

Fig. 1(a-b): (a) RT-PCR analysis using RNA-derived cDNA of the ORF0 fragment of *Sugarcane yellow leaf virus* from mature leaves and (b) Seedling of five sugarcane cultivars. H73-6110 and H87-4094 are susceptible, H65-7052 intermediately susceptible and H78-4153, H87-4319 resistant cultivars. Upper panel amplification products of ORF0 (360 bp); lower panel amplification product of 25S rRNA (108 bp) showing integrity of cDNA. The PCR products were electrophoresed on 1% agarose gel and visualized under UV light after staining with ethidium bromide. M: DNA molecular size markers (Fermentas, St. Leon Rot, Germany): O'GeneRuler^{Tm} Express 1 kb in upper panels of (a) and (b), 50 bp in lower panels of (a) and (b)

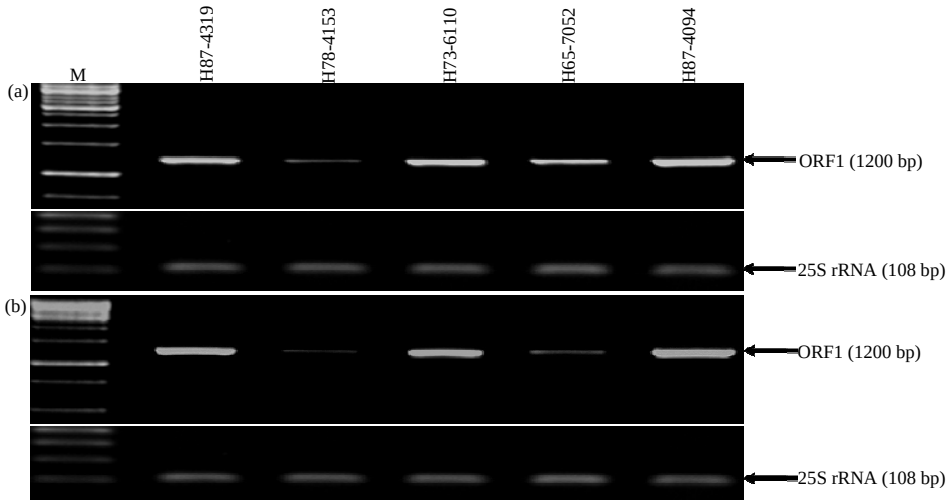

Fig. 2(a-b): (a) RT-PCR analysis using RNA-derived cDNA of the ORF1 fragment of *Sugarcane yellow leaf virus* from mature leaves and (b) Seedling of five sugarcane cultivars. H73-6110 and H87-4094 are susceptible, H65-7052 intermediately susceptible and H78-4153, H87-4319 resistant cultivars. Upper panel amplification products of ORF1 (1200 bp); lower panel amplification product of 25S rRNA (108 bp) showing integrity of cDNA. The PCR products were electrophoresed on 1% agarose gel and visualized under UV light after staining with ethidium bromide. M: DNA molecular size markers (Fermentas, St. Leon Rot, Germany): O'GeneRuler^{Tm} Express 1 kb in upper panels of (a) and (b), 50 bp in lower panels of (a) and (b)

Quantitative RT-PCR was employed to obtain a measure of transcript abundance for ORF0 and ORF1 related genes in all cultivars for quantitative comparison. The yield of transcripts for the two ORFs gives a variation in the viral accumulation in both tissues of mature leaves and seedling (Fig. 3a, b). The major amounts of PCR products were

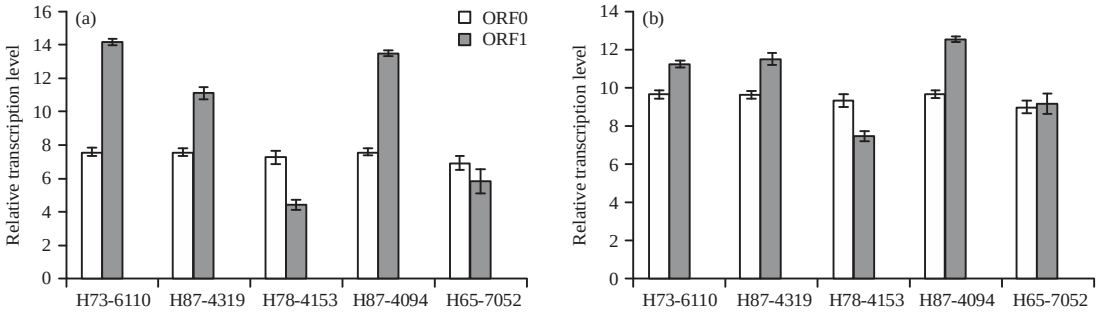

Fig. 3(a-b): Transcript levels of ORF0 and ORF1 genes in (a) Seedling tissues and (b) Mature leaves for Hawaiian sugarcane cultivars H73-6110 and H87-4094 (susceptible), H87-4319 and H78-4153 (resistant) and H65-7052 (intermediately susceptible). Transcript amounts were quantified by qPCR and related to GAPDH contents. Data are the means of 9 values corresponding to three biological (RNA extractions) replicates and three technical (PCR amplifications) replicates per sugarcane cultivar. The bar on top of each column represents the standard error of the mean

Table 2: Maximum likelihood estimation parameters used to describe substitution patterns in ORFs 0 and 1

Genes	ML estimate of substitution matrix[1]					Model[2]	Parameters[3]	BIC[4]	AICc[5]	R[6]	+G[7]	lnL[8]	Positions[9]
ORF0	-	A	T/U	C	G	K2+G	89	7310.965	6562.541	2.93	0.3284	-3178.643	741
	A	-	3.18#	3.18#	18.64*								
	T/U	3.18#	-	18.64*	3.18#								
	C	3.18#	18.64*	-	3.18#								
	G	18.64*	3.18#	3.18#	-								
ORF1		A	T/U	C	G	K2+G	53	13611.302	13144.367	3.31	0.2407	-6504.381	1838
	A	-	2.90*	2.90*	19.20*								
	T/U	2.90#	-	19.20*	2.90#								
	C	2.90#	19.20*	-	2.90#								
	G	19.20*	2.90#	2.90#	-								

1: Maximum likelihood estimate of substitution matrix, each entry is the probability of substitution (r) from one base (row) to another base (column). *Rates of different transitional substitutions, #Those of transversionsal substitutions, 2: Model with the lowest BIC scores (Bayesian information criterion) is considered to describe the substitution pattern the best, 3: No. of parameters (including branch lengths) used in the estimation, 4: Bayesian information criterion, 5: Akaike information criterion corrected, 6: The estimated transition/transversion bias, 7: Non-uniformity of evolutionary rates among sites may be modeled by using a discrete gamma distribution (+G) with 5 rate categories, 8: The maximum log likelihood estimation for automatically computed tree topology, 9: No. of positions in the final dataset

obtained for ORF1 in the seedling tissues of susceptible cultivars H73-6110 and H87-4094, while were lowest in resistant and intermediately susceptible cultivars H78-4153 and H65-7052, respectively. No significant differences were found between the cultivars for the ORF0 transcripts in mature leaves and immature tissues.

Amino acid sequence similarity comparison and maximum likelihood estimate of substitution matrix and transition/transversion bias: Amino acid sequence (available in the databases GenBank) similarity of ORFs 0 and 1 varied between 69-99 and 73-99%, respectively among SCYLV isolates.

Nucleotide substitution patterns and rates of ORFs 0 and 1 were estimated using the Kimura 2-parameter model (Table 2). For ORF0, a discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories [+G], parameter = 0.3284). Therefore, the general model formula for ORF0 was K2+G. Rates of different transitional substitutions had a probability of 18.64, while those of transversional substitutions had a probability of 3.18. The estimated transition/transversion bias (R) was 2.93. Substitution pattern and rates for ORF1 were also estimated using the model K2+G (+G = 0.2407). Rates of different transitional substitutions had a probability of 19.20, while those of transversional substitutions had a probability of 2.90. The estimated transition/transversion bias (R) was 3.31. A total of 741 and 1838 codon positions were identified in the final dataset for ORF0 and ORF1, respectively.

Inference about recombination and phylogeny: The TOPALi v2.5 program utilizing DSS statistics (window size: 500, step size: 10) was used to identify likely recombination events located in ORFs 0 and 1. The analysis exposed 2 and 5 major peaks which strongly supported the incidence of

recombination in aligned sequences of ORFs 0 in positions 310 and 435 and ORF1 in positions 380, 430, 549, 1420 and 1641 (Fig. 4a, b).

In order to confirm detection of possible recombinants, another program, RECCO was used and provided ORFs 0 and 1 (Fig. 5a, b) which displayed breakpoints represented by downward peaks in the dataset. The analysis yielded recombining fragments located in the intervals 217-275, 416-496 and 476-519 in ORF0's sequences (Fig. 5a) for the accessions HQ245348, HQ245359 and HQ245353, respectively. However, ORF1 (Fig. 5b) was more subject for recombination, RECCO demonstrated that recombination sites occurred in the intervals 457-476 and 519-573 in the sequence of the accession HQ245318; 637-669, 1039-1194, 1278-1290 and 1354-1442 in the sequence of the accession NC_000874; 1092-1107 and 1494-1585 in the sequence of the accession GU570006 and 1416-1445 in the sequence of the accession GU570007. Twenty four accessions were shown as putative recombinants. The results provided by RECCO algorithm for ORF1 were in agreement with RDP 4.1 algorithm only for isolates NC_000874, GU570006, GU570007 and HQ245318 (Table 3).

In inferring phylogenies, reconstructed trees showed that, in ORF0, all 45 available accessions in the data banks segregated into three major groups. Although group I encompassed genotypes CHN1, CHN2, BRA, PER and HAW, group II was composed of REU and CHN1 and group III was composed of CUB and IND (Fig. 6). In contrast, the 27 accessions of ORF1 was split into two major clusters (Fig. 7). In ORF1, group II was composed of genotypes REU and IND while group I was composed of the remaining seven genotypes, i.e., CHN1, CHN2, CHN3, BRA, PER, CUB and HAW (Fig. 7).

DISCUSSION

The tested Hawaiian cultivars of sugarcane contained SCYLV with different accumulation levels of virus titer including the resistant cultivars (H78-4153 and H87-4319) and the intermediately susceptible cultivar (H65-7052) which previously had been thought to be virus-free based on tissue blot immunoassay[16]. Rresults for studying the expression of two genes (ORF0 and ORF1) of SCYLV genome with RT-PCR showed that two Hawaiian cultivars previously had been considered to be uninfected and "Resistant[16]" were in fact infected with SCYLV (Fig. 1, 2). A study conducted in Hawaii using a tissue blot immunoassay shown that all plants were infected in some cultivars[16]. These cultivars were named SCYLV "Susceptible". On the other hand, some other cultivars seemed to be completely free of SCYLV and were called SCYLV "Resistant[16]". The SCYLV was never detected in these "Resistant" cultivars even when grown with close proximity to infected cultivars or when infested with viruliferous aphids[33]. The difference in qRT-PCR amplification among these resistant and highly susceptible cultivars was in the range of 5 fold (Fig. 3). Intriguingly, these resistant cultivars occasionally showed yellowing symptoms and it was always confusing if

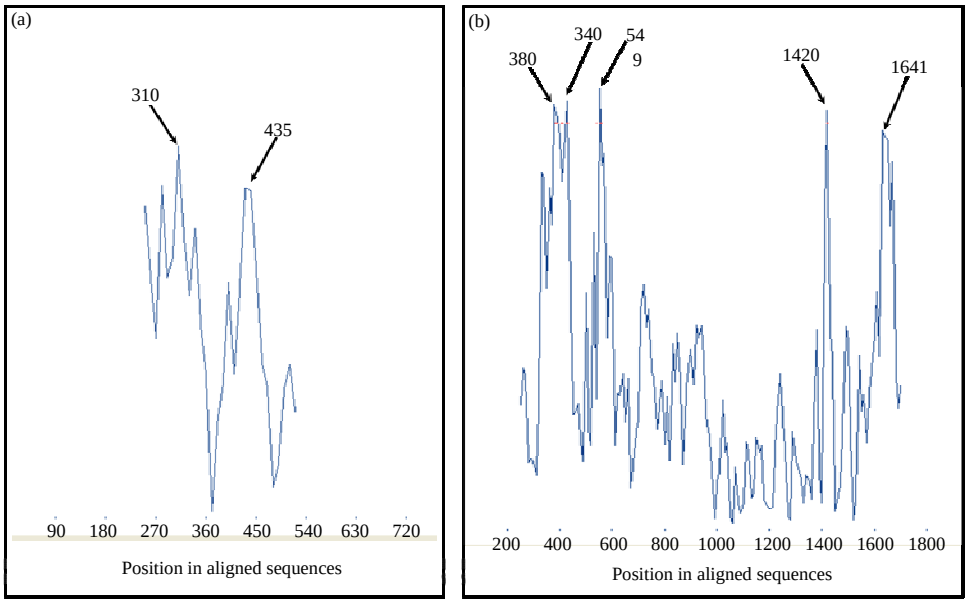

Fig. 4(a-b): Graph displaying potential breakpoint of recombination in (a) ORF0 and (b) ORF1 represented by a peak calculated by difference of sums of squares (DSS) statistics produced by using TOPALi v2.5 software

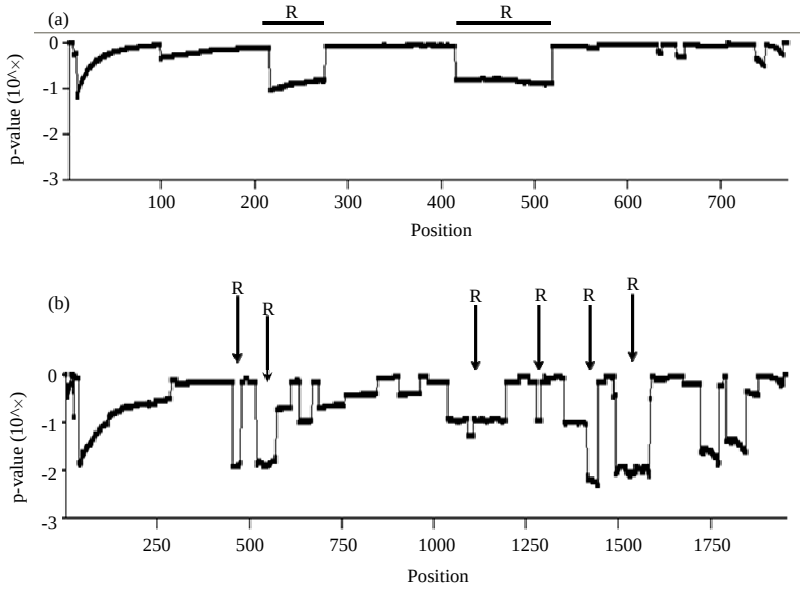

Fig. 5(a-b): Graph displaying potential breakpoint of recombination in (a) ORF0 and (b) ORF1 represented by a downward peaks determined by RECCO algorithm using cost optimization, R: Recombination site produced by using RECCO software

Table 3: Determination of inferred recombination events in the ORF1 sequences

Recombination pattern				p-value						
Recombinant isolate	Position in the sequence	Position in the alignment	Putative parental (major×minor)	R	G	B	M	C	S	3S
NC_000874	71-634	71-634	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-
KF477092	73-1096	73-1096	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-
KF477092	73-1096	73-1096	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-
GU570006	1875-end	1928-end	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-
GU570007	1787-end	1787-end	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-
GU570008	1878-end	1928-end	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-
GU327735	70-528	70-528	AM072756×Unknown	-	3.284×10^{-2}	-	1.310×10^{-3}	1.971×10^{-2}	7.484×10^{-33}	-
GU190159	Whole sequence	Whole sequence	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-
JF925152	46-504	70-528	AM072756×HQ245316	-	3.284×10^{-2}	-	1.310×10^{-3}	1.971×10^{-2}	7.484×10^{-33}	-
JF925153	46-504	70-528	AM072756×Unknown	-	3.284×10^{-2}	-	1.310×10^{-3}	1.971×10^{-2}	7.484×10^{-33}	-
JF925154	46-504	70-528	AM072756×Unknown	-	3.284×10^{-2}	-	1.310×10^{-3}	1.971×10^{-2}	7.484×10^{-33}	-
JF925155	46-504	70-528	AM072756×HQ245316	-	3.284×10^{-2}	-	1.310×10^{-3}	1.971×10^{-2}	7.484×10^{-33}	-
HQ245316	1921-end	1923-end	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-
HQ245317	1813-end	1813-end	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-
HQ245318	70-528	70-528	AM072756×HQ245316	-	3.284×10^{-2}	-	1.310×10^{-3}	1.971×10^{-2}	7.484×10^{-33}	-
HQ245319	142-634	142-634	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-
HQ245320	1921-end	1923-end	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-
HQ245321	1921-end	1923-end	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-
AM072751	103-617	603-617	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-
AM072752	1932-end	1934-end	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-
AM072753	1932-end	1934-end	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-
AM072754	68-604	68-604	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-
AM072755	68-641	68-641	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-
AM072756	73-604	73-604	Unknown×AM072750	-	3.319×10^{-2}	-	1.719×10^{-3}	>1	1.776×10^{-38}	-

The suite of recombination detection programs used for the detection of recombination events and the corresponding average p-values for each event were R: RDP, G: GenConv, B: Bootscan, M: MaxChi, C: Chimaera, S: Siscan, 3S: 3SEQ. 'Minor' and 'Major' parents refer to the parental isolates contributing the smaller and larger fractions of the recombinant's sequence, respectively, -: No recombination event detected

these symptoms were due to viral infection or were caused by physical damage such as broken midribs or severed stalks[33]. ORF0 revealed less variation in accumulation of virus titer compared with ORF1 (Fig. 3). The highest transcript gene was

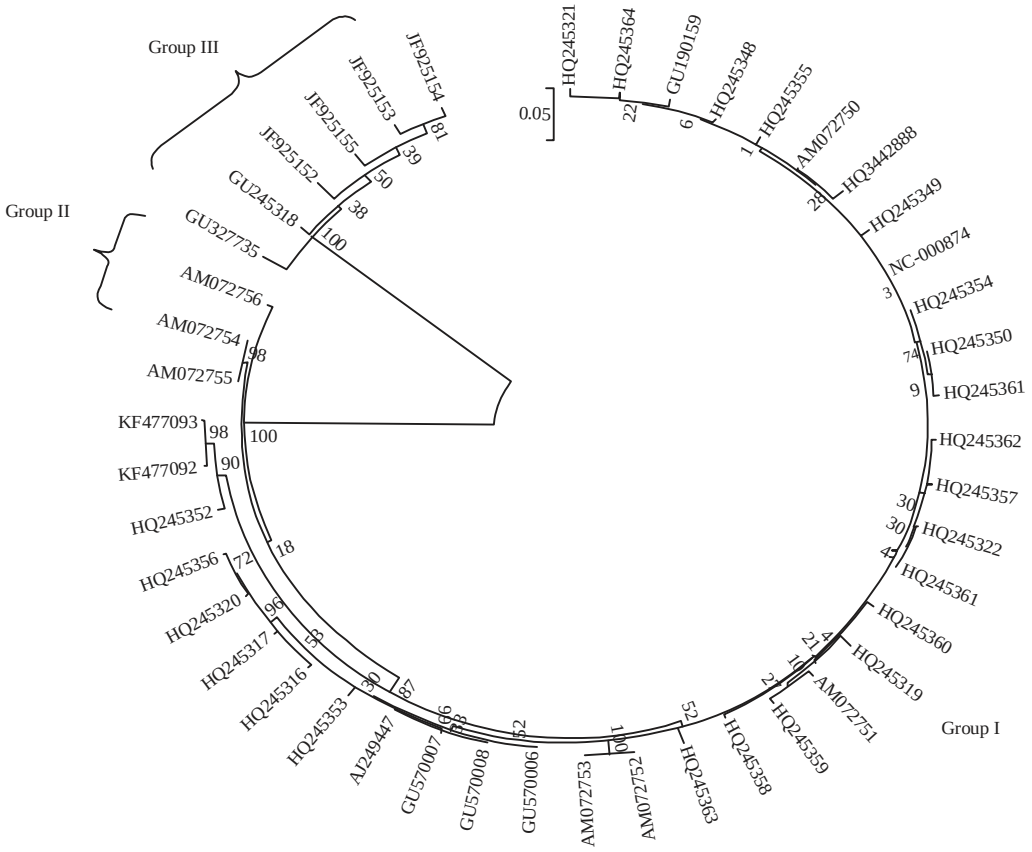

Fig. 6: Circle form of a phylogenetic tree produced by the ML algorithm under assumption of the Model K2+G an option of MEGA6 software. The ORF0 sequences of 45 SCYLV isolates distributed into three main clusters. The numbers above the branches indicate the bootstrap confidence value. The scale bar shows the number of substitution per nucleotide

ORF1 for cultivars H73-6110 and H87-4094 at the seedling stage; however, ORF1 was the most variant one in the seedling and mature tissues. Therefore, the diversity and changes in the genetic structure of plant virus populations are important aspects in plant pathology and may be highly relevant for developing strategies to control virus-induced diseases[34]. Determining the structure and function of a novel protein is a foundation of many aspects of modern biology. Excluding the structural proteins, not much is known about the functions of the putative products of the ORFs present in the SCYLV's genome as well as *Luteoviruses* in general. Which ORF encodes for the viral protein genome-linked (VPg) is still unclear and regarding the nonstructural genes, only the putative RNA-dependent RNA polymerase (by sequence comparison) and a putative movement protein have been identified[7]. Even though the translation product (p28-32) of ORF0 has been suggested to play a role in host recognition, the role of this ORF of *Luteoviruses* in the viral infection cycle has not been resolved[6].

To date, there is increasing evidence for a role for recombination in determining genetic diversity in some RNA viruses, mainly the *Retroviruses* and a variety of positive-sense RNA viruses including *Potyviruses*[35], *Luteoviruses*[36], *Nepoviruses*[37], *Closteroviruses*[38], *Cucumoviruses* and *Bromoviruses*[39,40] and most extraordinarily the *Caulimoviruses* in which the rate of recombination per base surpasses that of mutation[41]. The RNA recombination can be a proficient tool for viruses to repair viral genomes, therefore contributing to virus fitness. Moreover, it could play a role in the development of subviral RNAs that include defective interfering (DI) RNAs associated with many animal and plant viruses[42] as well as with chimeric satellite RNAs[43]. The indication for recombination, not only between related viruses but also among indistinctly related viruses and even with host RNAs, suggests that plant viruses test recombination with any genetic material available. In this study, it was demonstrated that ORFs 0 and 1 were subject to recombination. Interestingly, recombination events in ORF0 was detected with China isolates (HQ245348, HQ245359 and HQ245353) and not in Hawaiian isolates, though they are belonging to BRA-PRA genotype[13,15]. Thus the transcription level of ORF0 gene was consistent in Hawaiian isolates either

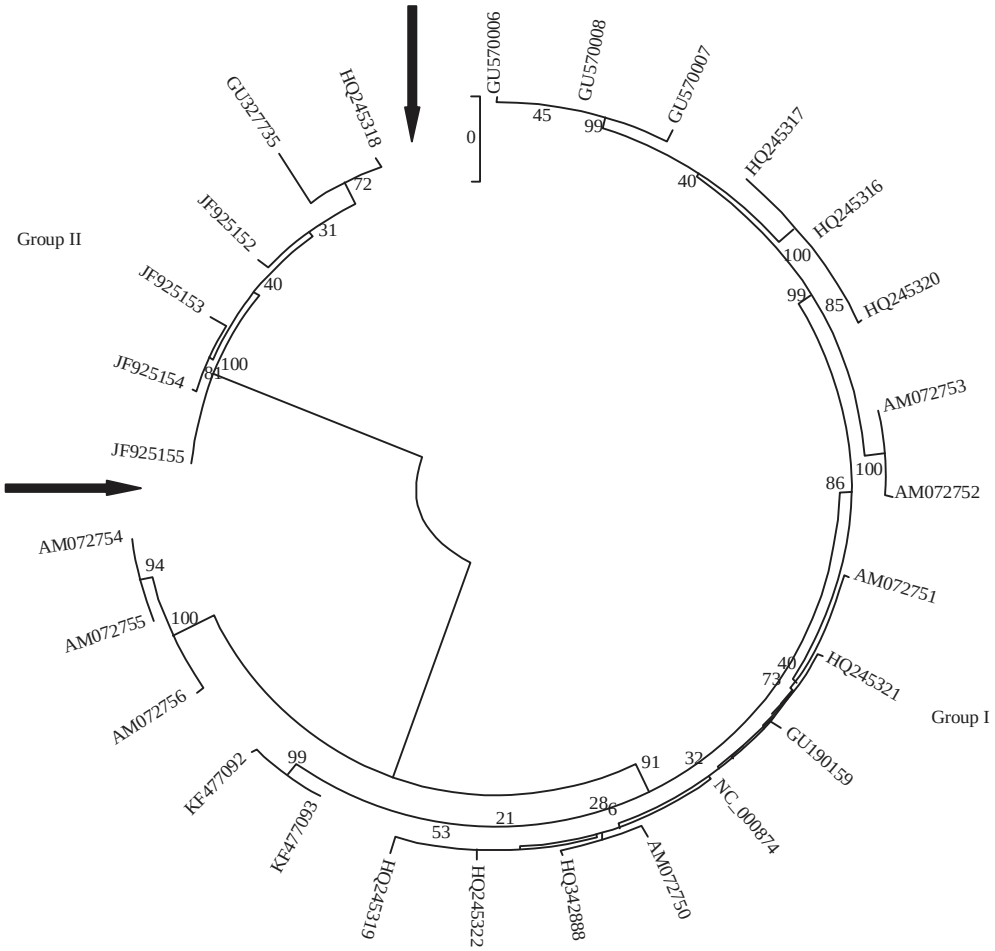

Fig. 7: Circle form of a phylogenetic tree produced by the ML algorithm under assumption of the Model K2+G an option of MEGA6 software. The ORF1 sequences of 27 SCYLV isolates distributed into two main clusters. The numbers above the branches indicate the bootstrap confidence value. The scale bar shows the number of substitution per nucleotide

in mature or sink tissues (Fig. 1, 3). It has been stated that the ORF0 (P0) play a critical role in virus biology, as P0 is an RNA silencing suppressor associated with virus pathogenicity[3,44].

The ORF0 was previously proposed as a sensitive and reliable diagnostic segment to discriminate among SCYLV genotypes[34,45] and to investigate the phylogeny of viral genotypes[44]. Amplification of complete ORF0 coding genes from different SCYLV isolates may facilitate the identification of new genotypes for which the full genome sequence is not available[15].

The ORF1 was more subject for recombination (Fig. 5b), RECCO demonstrated that recombination events were occurred in the sequence of the accessions: HQ245318, NC_000874, GU570006 and GU570007. The recombination analysis of 27 complete genomic sequences of SCYLV isolates reported worldwide revealed the presence of common potential recombination events, supporting previous reports that recombination is a dominant feature of SCYLV evolution, as in other RNA viruses[8,46,47]. Thus, environmental and host effects are likely to impact the rate of RNA recombination, in addition to the better-characterized viral factors[8]. In this study, Hawaiian isolates were significant recombinants in ORF1, but it was hardly to detect any recombination event in ORF0. The current findings contribute to understanding that recombination events may play an essential role in generating genome diversity. Because of no single phylogenetic tree can illustrate the evolutionary relatedness that exists between viruses because of recombination[48]. It is reported that the topology of SCYLV phylogenetic tree will be varied depending on the ORFs region (gene or the whole genome)[13,15]. This difference in clustering might be due to the fact that the SCYLV genes used for phylogenetic analysis have experienced different evolutionary histories[49].

It has been reported that RNA recombination is supposed to rescue viral genomes by repairing mutation errors in essential viral genes or in structures that could be introduced

during RNA replication[50,51]. Moreover, Worobey and Holmes[48] reported that RNA recombination enables exchange of genetic material between the same or similar viruses as well as between distinctly different viruses. Recently, ElSayed *et al*.[8] suggested that RNA recombination is linked to virus replication and that it occurs by a copy-choice mechanism. Also, they reported that Inter-species recombination has frequently occurred in the evolution of members of the *Luteoviridae*. The RNA recombination events probably produced the divergence observed between members of the genera *Polerovirus* and *Luteovirus*[52]. It is well known that mutation, recombination and purifying selection are important forces driving the evolution of SCYLV[45].

SIGNIFICANCE STATEMENTS

Recombination is the dominant feature of *Poleroviruses* as well as of SCYLV. Therefore, the present study aimed to enrich our understanding regarding to recombination of SCYLV. Analyses accomplished herein were based on the sequences of the ORF0 and ORF1 coding genes. These results showed that SCYLV contain potential recombination signals in ORFs 0 and 1. This recombination plays an important role in the genetic diversity among SCYLV populations.

CONCLUSION

Expression of ORF0 and 1 were varied among infected plants but overall expression of ORF1 was higher than ORF0. Possible recombination events located in the two ORFs were identified using TOPALi (v2.5), RECCO and RDP software's. These results showed strong presence of recombination in aligned sequences of ORF0 and ORF1 when TOPALi and RECCO programs were used. Two major forces, recombination and positive selection, drive the molecular evolution of viruses. We therefore, suggested further study on selective constraints exerting on CP gene might be a source of information to elucidate this.

ACKNOWLEDGMENTS

The researchers are grateful to World University Service, Germany for the gift of Thermal Cycler PCR and Gel imaging. This study was supported by the research fund from Zagazig University, Egypt and Centre for International Migration and Development, Germany.

REFERENCES

1. Moonan, F. and T.E. Mirkov, 2002. Analyses of genotypic diversity among North, South and Central American isolates of *sugarcane yellow leaf virus*: Evidence for colombian origins and for intraspecific spatial phylogenetic variation. J. Virol., 76: 1339-1348.
2. Smith, G.R., Z. Borg, B.E.L. Lockhart, K.S. Braithwaite and M.J. Gibbs, 2000. Sugarcane yellow leaf virus: A novel member of the *Luteoviridae* that probably arose by inter-species recombination. J. Gen. Virol., 81: 1865-1869.
3. Mangwende, T., M.L. Wang, W. Borth, J. Hu, P.H. Moore, T.E. Mirkov and H.H. Albert, 2009. The P0 gene of *Sugarcane yellow leaf virus* encodes an RNA silencing suppressor with unique activities. Virology, 384: 38-50.
4. Sadowy, E., A. Maasen, M. Juszczuk, C. David, W. Zagorski-Ostoja, B. Gronenborn and M.D. Hulanicka, 2001. The ORF0 product of *Potato leafroll* virus is indispensable for virus accumulation. J. Gen. Virol., 82: 1529-1532.
5. ElSayed, A.I., M. Boulila, E. Komor and Y.J. Zhu, 2012. Putative recombination signature and significance of insertion/deletion events in the RNA-dependent RNA polymerase coding region of *Sugarcane yellow leaf virus*. Biochimie, 94: 1764-1772.
6. Veidt, I., S.E. Bouzoubaa, R.M. Leiser, V.A. Ziegler-Graff, H. Guilley, K. Richards and G. Jonard, 1992. Synthesis of full-length transcripts of beet western yellows virus RNA: Messenger properties and biological activity in protoplasts. Virology, 186: 192-200.
7. Van Der Wilk, F., P. Houterman, J. Molthoff, F. Hans and B. Dekker *et al*., 1997. Expression of the potato leafroll virus ORF0 induces viral-disease-like symptoms in transgenic potato plants. Mol. Plant Microbe Interact., 10: 153-159.
8. ElSayed, A.I., E. Komor, M. Boulila, R. Viswanathan and D.C. Odero, 2015. Biology and management of sugarcane yellow leaf virus: An historical overview. Arch. Virol., 160: 2921-2934.
9. Abu Ahmad, Y., L. Rassaby, M. Royer, Z. Borg and K.S. Braithwaite *et al*., 2006. Yellow leaf of sugarcane is caused by at least three different genotypes of sugarcane yellow leaf virus, one of which predominates on the Island of Reunion. Arch. Virol., 151: 1355-1371.
10. Maia, I.G., M.C. Goncalves, P. Arruda and J. Vega, 2000. Molecular evidence that sugarcane yellow leaf virus (ScYLV) is amember of the *Luteoviridae* family. Arch. Virol., 145: 1009-1019.
11. Moonan, F., J. Molina and T.E. Mirkov, 2000. *Sugarcane yellow leaf virus*: An emerging virus that has evolved by recombination between luteoviral and poleroviral ancestors. Virology, 269: 156-171.

12. Chinnaraja, C., R. Viswanathan, R. Karuppaiah, K. Bagyalakshmi, P. Malathi and B. Parameswari, 2013. Complete genome characterization of *Sugarcane yellow leaf virus* from India: Evidence for RNA recombination. Eur. J. Plant Pathol., 135: 335-349.
13. ElSayed, A.I., M. Boulila and P. Rott, 2014. Molecular evolutionary history of *Sugarcane yellow leaf virus* based on sequence analysis of RNA-dependent RNA polymerase and putative aphid transmission factor-coding genes. J. Mol. Evol., 78: 349-365.
14. ElSayed, A.I., A.R. Weig and E. Komor, 2011. Molecular characterization of Hawaiian *Sugarcane yellow leaf virus* genotypes and their phylogenetic relationship to strains from other sugarcane-growing countries. Eur. J. Plant Pathol., 129: 399-412.
15. Lin, Y.H., S.J. Gao, M.B. Damaj, H.Y. Fu, R.K. Chen and T.E. Mirkov, 2014. Genome characterization of sugarcane yellow leaf virus from China reveals a novel recombinant genotype. Arch. Virol., 159: 1421-1429.
16. Schenck, S. and A.T. Lehrer, 2000. Factors affecting the transmission and spread of *Sugarcane yellow leaf virus*. Plant Dis., 84: 1085-1088.
17. Ruijter, J.M., C. Ramakers, W.M.H. Hoogaars, Y. Karlen, O. Bakker, M.J.B. van den Hoff and A.F.M. Moorman, 2009. Amplification efficiency: Linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res., Vol. 37, No. 6. 10.1093/nar/gkp045
18. Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res., 29: 2002-2007.
19. Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna and P.A. McGettigan *et al*., 2007. Clustal W and clustal X version 2.0. Bioinformatics, 23: 2947-2948.
20. Corpet, F., 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res., 16: 10881-10890.
21. Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar, 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol., 30: 2725-2729.
22. Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol., 16: 111-120.
23. Martin, D.P., P. Lemey, M. Lott, V. Moulton, D. Posada and P. Lefeuvre, 2010. RDP3: A flexible and fast computer program for analyzing recombination. Bioinformatics, 26: 2462-2463.
24. Martin, D.P., B. Murrell, M. Golden, A. Khoosal and B. Muhire, 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 10.1093/ve/vev003.
25. Maydt, J. and T. Lengauer, 2006. Recco: Recombination analysis using cost optimization. Bioinformatics, 22: 1064-1071.
26. Milne, I., F. Wright, G. Rowe, D.F. Marshall, D. Husmeier and G. McGuire, 2004. TOPALi: Software for automatic identification of recombinant sequences within DNA multiple alignments. Bioinformatics, 20: 1806-1807.
27. Martin, D.P., D. Posada, K.A. Crandall and C. Williamson, 2005. A modified Bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res. Hum. Retroviruses, 21: 98-102.
28. Padidam, M., S. Sawyer and C.M. Fauquet, 1999. Possible emergence of new geminiviruses by frequent recombination. Virology, 265: 218-225.
29. Smith, J.M., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol., 34: 126-129.
30. Posada, D. and K.A. Crandall, 2001. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. Proc. Nat. Acad. Sci. USA., 98: 13757-13762.
31. Gibbs, M.J., J.S. Armstrong and A.J. Gibbs, 2000. Sister-scanning: A Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics, 16: 573-582.
32. Boni, M.F., D. Posada and M.W. Feldman, 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics, 176: 1035-1047.
33. Lehrer, A.T. and E. Komor, 2008. Symptom expression of yellow leaf disease in sugarcane cultivars with different degrees of infection by *Sugarcane yellow leaf virus*. Plant Pathol., 57: 178-189.
34. ElSayed, A.I. and E. Komor, 2012. Investigation of ORF0 as a sensitive alternative diagnostic segment to detect *Sugarcane yellow leaf virus*. J. Gen. Plant Pathol., 78: 207-216.
35. Bousalem, M., S. Dallot, S. Fuji and K.T. Natsuaki, 2003. Origin, world-wide dispersion, bio-geographical diversification, radiation and recombination: An evolutionary history of *Yam Mild Mosaic Virus* (YMMV). Infect. Genet. Evol., 3: 189-206.
36. Gibbs, M., 1995. The Luteovirus Supergroup: Rampant Recombination and Persistent Partnerships. In: Molecular Basis of Virus Evolution, Gibbs, A.J., C.H. Callisher and F. Garcia-Arenal (Eds.). Cambridge University Press, Cambridge, UK., pp: 351-368.
37. Le Gall, O., M. Lanneau, T. Candresse and J. Dunez, 1995. The nucleotide sequence of the RNA-2 of an isolate of the English serotype of tomato black ring virus: RNA recombination in the history of nepoviruses. J. Gen. Virol., 76: 1279-1283.
38. Vives, M.C., L. Rubio, A. Sambade, T.E. Mirkov, P. Moreno and J. Guerri, 2005. Evidence of multiple recombination events between two RNA sequence variants within a *Citrus tristeza virus isolate*. Virology, 331: 232-237.
39. Bonnet, J., A. Fraile, S. Sacristan, J.M. Malpica and F. Garcia-Arenal, 2005. Role of recombination in the evolution of natural populations of *Cucumber mosaic virus*, a tripartite RNA plant virus. Virology, 332: 359-368.

40. Fraile, A., J.L. Alonso-Prados, M.A. Aranda, J.J. Bernal, J.M. Malpica and F. Garcia-Arenal, 1997. Genetic exchange by recombination or reassortment is infrequent in natural populations of a tripartite RNA plant virus. J. Virol., 71: 934-940.
41. Froissart, R., D. Roze, M. Uzest, L. Galibert, S. Blanc and Y. Michalakis, 2005. Recombination every day: Abundant recombination in a virus during a single multi-cellular host infection. PLoS Biol., Vol. 3. 10.1371/journal.pbio.0030089
42. Pathak, K.B. and P.D. Nagy, 2009. Defective interfering RNAs: Foes of viruses and friends of virologists. Viruses, 1: 895-919.
43. Cheng, C.P. and P.D. Nagy, 2003. Mechanism of RNA recombination in carmo- and tombusviruses: Evidence for template switching by the RNA-dependent RNA polymerase *in vitro*. J. Virol., 77: 12033-12047.
44. Lin, Y.H., S.H. Xiao, Y.H. Liu, J.S. Chen, H.Y. Fu, R.K. Chen and S.J. Gao, 2014. Molecular characterization and analysis of suppressing RNA silence of P0 protein encoded by *Sugarcane yellow leaf virus*. Scientia Agricultura Sinica, 47: 4627-4636.
45. Zhao, C.H., Y.H. Lin, Y.B. Pan, H.Y. Fu and R.K. Chen *et al*., 2016. Population structure and selection pressure analysis among *Sugarcane yellow leaf virus* isolates based on P0 and P1 sequences. Trop Plant Pathol., 41: 237-245.
46. Haenni, A.L., 2008. Virus Evolution and Taxonomy. In: Plant Virus Evolution, Roossinck, M.J. (Ed.). Springer, New York, USA., ISBN: 9783540757634.
47. Lian, S., J.S. Lee, W.K. Cho, J. Yu, M.K. Kim, H.S. Choi and K.H. Kim, 2013. Phylogenetic and recombination analysis of tomato spotted wilt virus. PLoS One, Vol. 8. 10.1371/journal.pone.0063380.
48. Worobey, M. and E.C. Holmes, 1999. Evolutionary aspects of recombination in RNA viruses. J. Gen. Virol., 80: 2535-2543.
49. Lin, Y.H., X.B. Wu, Y.H. Liu, S.H. Xiao and G.H. Zhou *et al*., 2015. Molecular detection, genetic diversity and distribution of *Sugarcane yellow leaf virus* genotypes in China. Trop. Plant Pathol., 40: 200-207.
50. Carpenter, C.D. and A.E. Simon, 1996. *In vivo* restoration of biologically active 3' ends of virus-associated RNAs by nonhomologous RNA recombination and replacement of a terminal motif. J. Virol., 70: 478-486.
51. Lai, M.M., 1992. RNA recombination in animal and plant viruses. Microbiol. Mol. Rev, 56: 61-79.
52. D'Arcy, C.J. and M. Mayo, 1997. Proposals for changes in luteovirus taxonomy and nomenclature. Arch. Virol., 142: 1285-1287.