

# Dynamics of genetic diversity among Indian *Sugarcane bacilliform virus* species and implications of associated recombination events in the virus

PK Janiga

ICAR Sugarcane Breeding Institute

K. Nithya

ICAR Sugarcane Breeding Institute

Rasappa Viswanathan (✉ [rasaviswanathan@yahoo.co.in](mailto:rasaviswanathan@yahoo.co.in))

Sugarcane Breeding Institute <https://orcid.org/0000-0002-7274-8144>

---

## Research Article

**Keywords:** Sugarcane bacilliform virus, RT/ RNase H, Phylogenetic analysis, Recombination, Population genetics

**Posted Date:** July 16th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-713719/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

*Sugarcane bacilliform virus* (SCBV), a plant pararetrovirus causing leaf fleck disease in sugarcane across the globe. Since the virus occurs throughout the sugarcane growing areas and germplasm in India, we have assessed the genetic divergences among the virus isolates from 104 isolates infecting germplasm and *Saccharum* hybrid varieties. With the evidence gathered from phylogenetic analysis and sequence demarcation tool, five novel genotypes are being proposed: SCBV-U, SCBV-V, SCBV-W, SCBV-X and SCBV-Y. SCBV-W: CBJ 46, due to the divergence in the nucleotide and protein sequence with existing isolates, established itself as a novel candidate in SCBV species. In comparison with the existing database of SCBV and conclusion from the present study, SCBV isolates from India exhibited maximum diversity in comparison with other regions. SCBV-U, SCBV-W, SCBV-X which turned out as potent recombinants along with other six recombinants in the study, directs at the plausible heterogeneity and genetic exchange happened within SCBV species over time which lead to the evolution of new variants. Neutrality tests indicate the existence of low-frequency polymorphism and Selection pressure of  $< 1$  pointed out at purifying selection. Codon usage bias acted as an efficient tool for identifying patterns in SCBV, in which non-randomness in mutational events might lead to the over expression of codon AGA. The current study on characterization and evidence of nucleotidal variation within SCBV species will lead to devising robust diagnostics of the virus in quarantine and improving the knowledge on the evolutionary changes in SCBV species.

# Introduction

Sugarcane is an important sugar and bioenergy crop grown under tropical and subtropical conditions. It meets nearly 80% of world's sugar requirement (<http://www.fao.org/faostat>). *Sugarcane bacilliform virus* (SCBV; genus *Badnavirus*, family: *Caulimo viridae*), a plant pararetrovirus that causes leaf fleck in sugarcane is reported from more than 20 sugarcane cultivating countries [10]. The virus is considered as an economically important pathogen and it limits exchange of sugarcane germplasm worldwide. SCBV was initially reported from Cuba in the commercial cv B34104 in 1985 [22]. The occurrence of the virus had been predominantly reported in numerous germplasm clones and cultivars from India, Australia, Brazil, China, Morocco, Mauritius, Guadeloupe, USA and other countries [1, 14, 19, 28, 32]. These viruses are naturally carried through infected planting materials and by sugarcane pink mealybug *Saccharicoccus sacchari* and grey mealybug (*Dysmicoccus boninsis*) in a semi-persistent manner [13]. Sugarcane is the major host for the virus and it also infects other grasses such as *Sorghum halepense*, *Brachiaria* spp., *Panicum maximum*, *Pennisetum* sp and *Rottboellia exaltata* [4, 14, 15, 28]. Unlike other Badnaviruses with a limited host range, SCBV is an exception in its wide host range, which has been proved to be transmitted to two different families, viz. *Poaceae* and *Musaceae* [6]. SCBV is reported widely from germplasm clones of *Saccharum officinarum* (noble cane) *S. barberi*, *S. robustum*, *S. spontaneum*, *S. sinensis* and *Pennisetum* sp, but can also infect commercial sugarcane (*Saccharum* interspecific hybrids) [15, 21, 31, 29].

Symptomatology of SCBV infection varied among germplasm and hybrid varieties [31]. Rao *et al.* [21] reported intensity of SCBV infection from mild to severe chlorosis with 28 different hybrids. Yellowish flecking and interveinal chlorotic streaks were also found among them. Later in a comprehensive study Viswanathan *et al.* [30] revealed that the symptoms were typical in *S. officinarum* and less common in *S. barberi*, *S. sinense*, *S. robustum* and interspecific hybrids. In *Saccharum* spp. clones, common symptoms comprise varying degrees of chlorotic stripes, stunted growth, severe chlorotic mottling, and pronounced fleck. The symptoms started to appear prominently from tillering phase of the crop, and the severity continued to increase with ageing. However, in hybrid clones, the symptoms appeared as intense flecks to mild mottle at the distal portion of the leaf lamina gradually progressed towards the proximal region. The chlorotic flecks turned yellow and subsequently leading to the reddening of entire leaf lamina [30].

SCBV has a non-covalently closed double-stranded DNA genome of 7.3–7.8 kilo base pair (kb) in size and replicates via a virus-encoded reverse transcriptase (RT) [5, 9]. The genome is not circular because of site-specific discontinuities in the negative sense and positive sense strand. SCBV owns a typical badnavirus genome organization comprised of three open reading frames (ORFs). ORFs 1 and 2 encode for a small protein of unknown functions while the ORF3 encodes for functional proteins viz. movement protein, coat protein, aspartic protease, reverse transcriptase (RT) and ribonuclease H (RNase H) protein [10, 19, 24].

The RT/RNase H coding region is a common taxonomic marker for species demarcation within the genus *Badnavirus*. The demarcation criterion is > 20% nucleotidal variation in the RT/RNase H region, proposed by the International Committee on Taxonomy of Viruses (ICTV) [9]. Currently, four species of SCBV were recognized such as *Sugarcane bacilliform IM virus* (SCB IMV), *Sugarcane bacilliform MO virus* (SCBMOV), *Sugarcane bacilliform Guadeloupe A virus* and *Sugarcane bacilliform Guadeloupe D virus* by ICTV as different species under *Badnavirus* [9, 19]. Currently, 20 genotypes (SCBV-A to SCBV-T) of SCBV have been reported worldwide. Seven genotypes viz, SCBV-E, H, I, J, K, L and M [1, 32]; five from France, SCBV A, B, C, D and G [19] and five from Brazilian germplasm SCBV A, C, F, M and H [6].

In recent years, extensive sequence information was generated from various sugarcane growing provinces across the globe. Our earlier studies revealed widespread occurrences of the virus and genomic diversity in India, the second major sugarcane growing country in the world [10, 30]. Hence, further studies were taken up to address the genetic variability among SCBV to develop precise diagnostic assays for disease-free clones. This study with 104 new SCBV isolates from germplasm and hybrid varieties brought out new information on the prevalence of enormous genetic variation among SCBV population in India for the first time and involvement of recombination events associated with genomic variation in the virus population.

## Materials And Methods

### Leaf samples

Detailed surveys were made to collect leaf fleck suspected sugarcane leaf samples from various parts of India and in *Saccharum* germplasm during 2019–2020. Totally, samples from 125 germplasm clones and 233 cultivated varieties were collected from the states of Tamil Nadu, Kerala, Karnataka, Maharashtra, Andaman and Nicobar islands (Table 1). The germplasm clones comprised of *S. officinarum*, *S. spontaneum*, *S. barberi*, *S. sinense*, *S. robustum*, *Erianthus arundinaceous*, *E. bengalensis*, *Sclerostachya* sp, *Narenga* sp and hybrid cultivars include *Saccharum* hybrid varieties and *Saccharum* interspecific hybrid varieties and exotic clones. All samples were stored at -80°C until DNA extraction. The leaf samples were ground in liquid nitrogen and DNA was isolated using CTAB buffer, as reported earlier [10]. The quality of the DNA sample was quantitatively checked in a Nanodrop spectrophotometer (Thermo scientific, USA) and qualitatively by agarose gel (0.8%) electrophoresis.

Table 1

Occurrence of SCBV in the cultivated varieties and germplasm clones from sugarcane growing areas in India, established with SCBV 794 primer.

Samples	Total number of samples	SCBV positive samples	Percentage of detection
<i>S. officinarum</i>	56	24	48 %
<i>S. spontaneum</i>	40	20	
<i>S. barberi</i>	5	3	
<i>S. sinense</i>	6	2	
<i>S. robustum</i>	5	2	
<i>E. arundinaceous</i>	10	6	
<i>E. bengalensis</i>	1	1	
<i>Sclerostachya</i> spp.	1	1	
<i>Narenga</i> spp.	1	1	
<i>Saccharum</i> hybrids	218	126	57.08 %
<i>Saccharum</i> interspecific hybrid	9	3	
Exotic clone	6	4	

## Primer Designing and PCR amplification

Whole-genome sequences available to date were retrieved from NCBI and were aligned using BioEdit 7.0.5.3. RT/RNase H region was selected as the region for PCR amplification based on the ICTV criteria for species demarcation. Failure of the established primers to identify various types of SCBV led to designing a new set of degenerate primers targeting the conserved motifs of ORF3 polyprotein-RT/RNaseH region. The primers SCBV 794 FP-5'GCRCCWGCAGT VTTYCARAGGAAGATG3' and SCBV 794 RP-5' CCAYCTGATCTCHGAAGGYTTRTG 3' have specifically amplified ~ 794bp fragment. About 300-

500ng of total genomic DNA was subjected to PCR using *Taq* polymerase (TaKaRa, Japan) by following the optimized assay conditions: an initial denaturation step at 94°C for 5 min, followed by 35 cycles of 94°C for 45s; 59°C for 45 s; 72°C for 45 s; and a final extension step at 72°C for 10 min. Amplified DNA was resolved on 1.5% agarose gels pre-stained with ethidium bromide and then visualized under a gel documentation system to verify their size specificity (GBox, Syngene). PCR-positive sample derived from a single sugarcane variety was considered as one isolate, hereafter will be referred to as an SCBV isolate.

### **Sanger di-deoxy sequencing and in silico sequence analysis**

The amplicons were purified in 15µl of nuclease free water using GenElute Gel Extraction kit (Sigma, USA) and sequenced by Sanger di-deoxy sequencing (Eurofins Scientific, Bangalore). Forward and reverse strands of the isolates were sequenced and the obtained sequences were aligned to form a contiguous sequence using the Cap contig assembly programme of Bioedit 7.0.5.3. The resulted contigues were used to query the National Centre for Biotechnology Information (NCBI) database ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) with the BLASTn search functions, where BLAST analysis provided match to viral sequences.

## **Phylogenetic profile analysis**

Contig sequences were used for inferring evolutionary relationship among SCBV species. Phylogenetic analysis was done using the nucleotide sequences of 603 bp RT/RNase H-coding regions delineated by the SCBV794 FP/RP primer binding sites. The analysis included 104 sequences from the present study, 23 SCBV genome sequences and one *Rice tungro bacilliform virus* sequence (RTBV Accession no: NC\_001914.1) (as out group) from the NCBI database. Twenty sequences from the NCBI represented 20 genotypes established (SCBVA-T) earlier. All the sequences were trimmed correspondingly to 603nt; aligned using ClustalW and the phylogenetic tree was constructed using the Maximum likelihood method with Tamura-Nei model [26] with MEGAX software [11]. The robustness of the bootstrap consensus tree was inferred from 1000 replicates.

## **Pairwise Sequence Alignment and Identity Calculation**

Recently pairwise-identity-based viral classification has proven highly useful and popular amongst viral taxonomists particularly with small viral genomes. Amongst pairwise-identity-based classification approaches, sequence demarcation tool - SDT 1.2 (available from <http://web.cbio.uct.ac.za/SDT>) software has been adopted for pairwise sequence alignment and identity calculation. The identity scores were calculated using ClustalW alignment approaches as  $1-(M/N)$  where M, the number of mismatching nucleotides and N, the total number of positions along the alignment at which neither sequence has a gap character [18]. We have included the previously established 20 genotypes/species of SCBV with five possible genotypes from the current study derived through phylogenetic analysis for identity calculation.

## **Outline of recombination events in SCBV**

Inter recombination among SCBV isolates was achieved through Recombination detection programme 4.39 (RDP4) [16] by implementing the following methods: RDP, GENECONV, Bootscan, Maxchi, Chimaera, Siscan and 3seq methods. The strong point of RDP4 is that it simultaneously uses various recombination

detection methods to detect and characterize the recombination events that are apparent within an aligned nucleotide sequence without any prior user indication of a non-recombinant reference sequence. A data file of 124 SCBV sequences was prepared in nexus format and potential recombination events detected using the above mentioned methods. RDP, 3Seq, GENECONV, Maxchi and Chimaera methods mainly examine variable nucleotide positions in triplets of sequences sampled from the alignment, while the Bootscan and Siscan methods examine all variable and conserved nucleotide positions. The first four methods comprise the detection of recombination signals in the alignment and the second involves inference of the number and characteristics of unique recombination events that have generated these signals. Recombination breakpoint hot-spots were established from the breakpoint distribution plot using the permutation-based test.

## **Assessment of population genetic parameters**

Population genetic analysis involves the genetic variation within a viral population driven by a myriad of factors including recombination, mutation rates, a pattern of selection and stochastic noise caused by random genetic drift in that order. The neutrality of the SCBV population was calculated using DNA sequence polymorphism (DnaSP) 5.10.01 software [12]. One hundred four isolates from the study and 20 genotypes established earlier were used for carrying out the parameter. Neutrality test was conducted for verifying the neutral mutation hypothesis by using an average number of nucleotidal difference in the genome and the number of segregation sites employing Tajima's D test [25] and Fu and Li's D [8].

Selection pressure on the coding region was assessed using the SLAC (Single-Likelihood Ancestor Counting) analysis using data monkey (<http://www.datamonkey.org/>) server by HKY85 nucleotide substitution bias model. Generally, SLAC uses maximum likelihood to infer ancestral characters for nucleotides across the phylogeny and then directly counts the number of synonymous and non-synonymous alterations at each site over evolutionary time. The ratio of non-synonymous substitution; dN (amino acid altering substitutions) and synonymous substitution; dS (substitutions that do not modify amino acids) (dN/dS) has been broadly used as an indicator of selection pressure ( $\omega = dN/dS$ ) [20].

## **Relative synonymous codon usage (RSCU) pattern in SCBV**

The codon usage pattern for all the established and proposed genotypes were summarized using Mega X software. RSCU value of a codon is defined as the ratio of its observed frequency to its expected frequency providing all codons for a particular amino acid are used equally. RSCU values < 1.0, 1.0, and > 1.0 correspond to negative codon usage bias, no bias, and positive bias, respectively [23].

## **Results**

### **PCR amplification of RT/RNase H fragments and sequence analysis**

Of the 358 sugarcane samples, 57% (133/233) of *Saccharum* hybrid varieties and 48% (60/125) germplasm clones were positive with SCBV794 primer, which gave an amplicons size of 794bp. Apparently, based on the amplicons image, the quantity of viral titre was found to be more in the cultivated varieties as compared to germplasm samples (Fig. 1). The primer detected the virus from a broad spectrum of infected sugarcane, mainly *S. officinarum*, *S. spontaneum*, *S. barberi*, *S. sinense*, *S. robustum*, *E. arundinaceous*, *E. bengalensis*, *Sclerostachya* spp, *Narenga* spp and various *Saccharum* hybrid varieties. Contigues derived from the partial sequencing of RT/RNaseH region of all SCBV isolates showed 80–99% similarity with existing SCBV sequences in NCBI. All the contigues derived from the present study were submitted to NCBI-Genbank (Supplementary Table 1).

## Phylogenetic profile analysis

Evolutionary analysis using the maximum likelihood method revealed the segregation of sequences into three major monophyletic groups where most of the isolates clustered in the third monophyletic group. Broadly, the phylogenetic tree (Fig. 2A) classified the isolates into 25 genotypes; five genotypes from the study SCBV-U, V, W, X and Y and 20 genotypes reported from all over the world (SCBVA-T). Fifty-nine isolates from the study were grouped into a separate cluster forming a new genotype SCBV-U (Fig. 2B). Another 17 isolates formed a branch of SCBV-L genotype (Fig. 2C). The isolate CBJ 46 showing < 88% similarity to the neighbouring N genotype SCBV-FJZZ3 from China, formed a distinct branch (SCBV-W) outside the 3rd monophyletic tree. Further, the isolate CBJ 46 is the only isolate that showed no similarity to any of the existing whole-genome sequences. The isolates from germplasm viz. *S. officinarum* clones (Bangadya, Saipan G, Baragua), *S. spontaneum* 81–095 and interspecific hybrids, ISH 1 and Cym 08-666 formed a separate cluster forming the novel genotype SCBV-V. Notably, *S. barberi* isolate MW548486 shared identity of 87.57% with SCBV-CHN2 (KM 214358.1) but grouped far from it while forming a novel subclade SCBV-X. Likewise, *Saccharum* hybrid Co 2001-13 from Karnataka state showed similarity to SCBV-BRU with 86.5% similarity in their nucleotide composition; yet formed phylogenetically distinct clade forming the new genotype SCBV-Y. In contrary, an isolate from Coimbatore (MW548472) from the same sugarcane hybrid formed the genotype SCBV-U from the study. Five genotypes that were already established in other sugarcane growing countries SCBV-G (France and China), SCBV-Q (China), SCBV-R (China), SCBV-S (China) and SCBV-T (China) were reported first time from India; SCBV-G: *Saccharum officinarum* Poona, SCBV-Q: CoC 24, *Saccharum spontaneum* IND 81–157, *Saccharum spontaneum* IND 81–086, SCBV-R: *Saccharum spontaneum* SES 954, *Saccharum spontaneum* IND 85–535, *Saccharum spontaneum* IND 81 – 003, *Saccharum spontaneum* IND 84–426, SCBV-S: *Saccharum spontaneum* 07-1488, *Saccharum spontaneum* IND 85–512, SCBV-T: *Saccharum spontaneum* IND 84–432 and *Saccharum officinarum* Khajuria, *S. officinarum* isolates MW548484 and MW548485 formed a cluster with SCBV-H genotype in the group 1. *S. officinarum* isolate MW645069 emerged as genotype E with 92.7% similarities to that of SCBV Iscam reported earlier from India [10]. SCBV-U, a novel genotype from the present study can be considered as the most frequently occurring genotype in India, especially in case of isolates from *Saccharum* hybrids and interspecific hybrids.

## Pairwise sequence alignment and identity calculation

Pairwise sequence identity among the 25 SCBV genotypes was established (Fig. 3). All the five possible new genotypes deduced from the phylogenetic analysis viz. SCBV: U-Y showed < 92% nucleotidal similarity to remaining species/genotypes, thus proves to be accepted as novel genotypes. Pairwise identity calculation was able to distinguish between the species and strain/genotype among the sequences. SCBV-W (isolate-CBJ 46) is a novel genotype with the least similarity (< 80%) to the other four new genotypes, but it needs further confirmation through complete genome analysis. SCBV: A - D from Guadeloupe showed the highest divergence with 56.7–61.7% followed by SCBV-M [21], SCBV-E (*Sugarcane bacilliform IM virus*), SCBV-H [10], SCBV-N [32] and SCBV-W (this study). SCBV G-H genotypes showed the most similarity to SCBV-F among all the ICTV approved species/strains. Five proposed genotypes proved the variation in the nucleotides usually present in the *Badnavirus* genus with percentage similarity of SCBV-U (59.2–86.4%), SCBV-V (59.2–91.7%), SCBV-X (57.8–86.9%) and SCBV-Y (56.1–88.4%). The results obtained from the sequence demarcation tool are in good harmony with the phylogenetic pattern.

## Distribution of variation across the amino acids in RT/RNase H region

In present study, 152 amino acid sequences retrieved from RT/RNase H region of ORF 3 functional protein of 25 genotypes were aligned (Fig. 4). Amino acids from 1-9th position "EEEHAEHL" in SCBV E-Y were replaced by "VQQHKEHLK" in SCBV A-D genotypes. A single amino acid change was found throughout the selected region of SCBV genotypes. Variants reported from countries like India, China and Australia showed similarity in protein sequences whereas more significant dissimilarity was found with Guadeloupe genotypes. The 10th amino acid (in the alignment) showed the maximum dissimilarity in the position with K, I, V, A, E, T and N. Resemblance of amino acids within the SCBV E-Y genotypes advocates the origin of these isolates from a common ancestor, which might have undergone continued exchange of gene fragments.

## Outline of recombination events in SCBV

RT/RNase H region of badnaviruses is considered as the species demarcation region with greater than 20% variability. Greater differences in nucleotides generally present in SCBV lead to the emergence of numerous genotypes, thus making recombination and genomic reassortment possible inside the genome. Exploring a dataset of 124 nucleotides with seven methods using RDP4 revealed the presence of inter-SCBV recombination. Out of 68 recombination events detected, a total of 9 events were found to be significant, where the p-value was 0.05 (Table 2). Fifty-eight isolates from the SCBV-U genotype showed trace evidences of recombination, suggested them as a recombinant genotype. The SCBV isolates CBJ 46, ISH 101 and SB Pathri from the current study with the proposed genotypes of SCBV-W, U and X respectively were established as recombinants. SCBV-W variant CBJ46 became a recombinant with SS IND 81 – 003 (genotype R) as a major parent and PR 1062 (genotype U) as a minor parent. SS IND 81 – 003 is a recombinant isolate from SCBV-BRU (major parent) and CB622 (minor parent). SCBV-CHN2 (genotype G) with SS IND 81–157 (genotype Q) contributed to the evolution of SB Pathri isolate (genotype X). SS IND 81 – 003 (genotype R) and SCBV Thiruvalla (genotype M) with a significant p-value



of  $1.15 \times 10^{-02}$  add to the recombinant ISH 101 (genotype U). SO Otaheite (genotype H) became recombinant with major fragment from SS IND 81–157 and minor fragment from SO BS Aubin. Exchange of genetic fragments took place between the following isolates- SCBV GD-YT2361 (genotype R), CBSe 95436 (genotype L) and CBC 24 (genotype Q) to form SCBV-R, SCBV-L and SCBV-Q genotypes. SCBV FJZZ3 (genotype N) reported from China is a recombinant from Q 70 (genotype U) and SCBV HN-GT1271 (genotype P). From the analysis, SCBV-U and SCBV-R genotypes played a key role as major or minor contributors of recombination events. The genetic exchange must have happened through germplasm materials across provinces/ countries that led to new variants, which is evident from the recombination plot.

Table 2

Recombination events detected using RDP4. Major parent and minor parent are the sequences which contributed to the formation of recombinant. R, G, B, M, C, S and 3s are acronyms used for the methods viz. RDP, GENECONV, BOOTSCAN, MAXCHI, CHIMERA, SISCAN and 3SEQ to find out intra SCBV recombination with a p- default value of 0.05

Event no:	Recombinant	Major parent	Minor parent	Breakpoints (nucleotide position)		Methods	p-value
				Start	End		
R1	CBJ 46	SS IND 81 – 003	PR 1062	30	208	MCS3s	$1.15 \times 10^{-02}$
R2	SS IND 81 – 003	SCBV BRU	CB 622	318	563	MCS	$1.16 \times 10^{-02}$
R3	SB Pathri	SCBV CHN2	SS IND 81–157	347	541	MCS3s	$1.58 \times 10^{-02}$
R4	SO Otaheite	SS IND 81–157	SO B S Aubin	149	333	RGMCS	$3.41 \times 10^{-02}$
R5	ISH 101	SS IND 81 – 003	SCBV Thiruvalla	185	321	MCS 3s	$1.150 \times 10^{-02}$
R6	CBSe 95436	CBC 24	SCBV GD-YT2361	31	148	RGCM	$2.29 \times 10^{-02}$
R7	SCBV MO	CB 2001-13	SCBV CHN2	266	486	MCS3s	$2.44 \times 10^{-02}$
R8	SCBV FJZZ3	Q 70	SCBV HN-GT1271	50	255	RGBMCS	$2.03 \times 10^{-02}$
R9	SCBV GD-YT2361	CBSe 95436	CBC 24	1	202	RGBMC	$2.91 \times 10^{-01}$

Furthermore, genomic fragments were derived from distinct parents who belong to different phylogenetic clusters. Breakpoint distribution plot of 124 isolates with 603nt sequence established 250–550 (position in the alignment) as a probable hot spot inside the RT/RNase H region (Fig. 5). Characterization of

recombination events and gene assortment within these genotypes of SCBV will offer insight into the evolution of the species over the generation. The breakpoint distribution plot aided in, finding out the frequent variation sites within the RT/RNase H region of the genome.

## Assessment of population genetic parameters

Neutrality tests of 124 SCBV populations with mutation site – 450 and segregation site – 260 gave a D value of -1.142 and – 0.966, respectively for Tajima's D test and Fu and Li's Test (Table 3). The average number of nucleotidal differences between these populations (K) was 54.769 and the nucleotide diversity (Pi) was 0.143. A negative neutrality test signified an excess of low-frequency polymorphisms indicating a selective swap/expansion of population after a current bottleneck. Selection constraints on the coding region over each nucleotide data were estimated with dN/dS ratio. A mean value of < 1 indicated a negative selection (Fig. 6). Out of 201 sites considered, SLAC found evidence of pervasiveness, a negative/purifying selection at 156 sites with a p-value threshold of 0.1.

Table 3  
Neutrality test and its significance in SCBV population was accessed using Tajima's D ; Fu and Li's test statistics.

<b>Input : 124 sequences</b>	<b>Tajima's D</b>	<b>Fu and Li's test</b>
Number of polymorphic segregating sites; <b>S</b>	260	260
Total number of mutations; <b>Eta</b>	450	450
<b>D value</b>	-1.142	-0.966

## Relative Synonymous Codon Usage (RSCU) in SCBV

Generally, the RSCU value of > 1.05 considered as overexpressed while < 0.4 is an under represented codon. Codon usage in SCBV biased towards AGA (Arginine); one of the most overexpressed ones in SCBV A-Y genotypes (RSCU value: 2.4–5.14). Following codons are also well expressed in SCBV viz., GGA (Glycine), CCA (Proline), GCA (Alanine), ACA (Threonine) and AUU (Isoleucine). UCA (Serine), deemed to be less expressed codon among genotypes with < 0.4 representation. RSCU value of codon CGC, CGU, CGA, CGG, which codes for arginine was null in most of the genotypes, may be due to the higher representation of AGA codon. UCG (serine) codon was found absent in all the genotypes except SCBV-U and SCBV-W. The discrepancy in the percentage of representation of codons in various genotypes was evident from the RSCU matrix (Table 4).

Table 4

Relative synonymous codon usage is given in parentheses following the codon frequency. All frequencies are averages over all taxa. Codon usage biases in twenty five genotypes of SCBV were used for tabulation.

Codon	RSCU	Codon	RSCU	Codon	RSCU	Codon	RSCU
UUU(F)	1.3	UCU(S)	0.67	UAU(Y)	1.36	UGU(C)	1.16
UUC(F)	0.7	UCC(S)	0.88	UAC(Y)	0.64	UGC(C)	0.84
UUA(L)	1.51	UCA(S)	1.61	UAA(*)	0	UGA(*)	0
UUG(L)	0.72	UCG(S)	0.11	UAG(*)	0	UGG(W)	1
CUU(L)	1.02	CCU(P)	0.9	CAU(H)	1.39	CGU(R)	0.09
CUC(L)	0.75	CCC(P)	0.98	CAC(H)	0.61	CGC(R)	0.06
CUA(L)	0.9	CCA(P)	1.83	CAA(Q)	0.82	CGA(R)	0.47
CUG(L)	1.1	CCG(P)	0.28	CAG(Q)	1.18	CGG(R)	0.09
AUU(I)	1.3	ACU(T)	1.17	AAU(N)	1.68	AGU(S)	1.1
AUC(I)	1.15	ACC(T)	0.61	AAC(N)	0.32	AGC(S)	1.63
AUA(I)	0.54	ACA(T)	2.05	AAA(K)	0.88	AGA(R)	4.35
AUG(M)	1	ACG(T)	0.16	AAG(K)	1.12	AGG(R)	0.93
GUU(V)	0.94	GCU(A)	1.13	GAU(D)	1.68	GGU(G)	0.81
GUC(V)	0.97	GCC(A)	0.55	GAC(D)	0.32	GGC(G)	0.42
GUA(V)	1.12	GCA(A)	2.28	GAA(E)	1.27	GGA(G)	1.97
GUG(V)	0.97	GCG(A)	0.04	GAG(E)	0.73	GGG(G)	0.79

## Discussion

The exchange of sugarcane germplasm for breeding and commercial production across the globe has played a vital role in expanding the sugar industry worldwide. SCBV is considered one of the frequently detected viruses in quarantine during germplasm exchange, thereby considered an economically important pathogen [2]. Surprisingly there are no procedures yet available to remove this virus from sugarcane materials. Apical meristem culture (AMC), commonly used to eliminate other viruses, is not regarded as an appropriate and efficient method, since this virus is expected to infect meristem tissues. Discarding sugarcane infected with this virus would result in a loss of 30–40% of the germplasm quarantine materials [7]. Hence adequate detection methods have to be developed to detect the virus from germplasm. The knowledge regarding the prevalence and genetic diversity of the virus is the prime objective for molecular diagnosis. The Institute maintains World Sugarcane Germplasm Collections at Kannur, Kerala, which is the world's largest germplasm collection and it comprises all the species of

*Saccharum*, different species of allied genera viz. *Erianthus*, *Narenga*, *Sclerostachya*, *Pennisetum*, foreign hybrids and Indian hybrids numbering 3373. Since the studies on the occurrence of SCBV in India are limited, the current work will showcase a comprehensive assessment of SCBV from sugarcane germplasm and cultivated varieties.

SCBV from one of the popular varieties Co 86032 collected during 2019–2020 from different fields Pune, Maharashtra state, Avinashi, Neelambur and Vedapatti, Tamil Nadu state and different generations tissue culture derived canes (T0, T1 and T3) showed greater similarity in nucleotide composition and clustered together in the new subclade SCBV-U. Likewise, SCBV detected from the hybrid cv Co 0212 collected from different fields viz., SBI-VPT, Indiyampalayam and Avinashi in Tamil Nadu state dispersed in the same genotype clade SCBV-U. Similarly samples from *Saccharum* hybrid cv PI 1110 from three different regions shared same genotype identity. The existence of same genotype from different fields points out that, the virus might have transmitted through true seed. Recently it is found that the virus is transmitted through true seed [3] and the virus isolate from the parental clone is carried to the progenies and maintained through generations of vegetative propagation. In contrast to this, SCBV isolate from the cv Co 2001-13 from Karnataka (MW5484708) segregated as a distinct clade SCBV-Y while its counterpart from Tamil Nadu MW548472 showed identity to SCBV-U genotype. This may probably due to possible infection of a different isolate through mealy bug vector in a new location. In general, the hybrid varieties of sugarcane showed a greater susceptibility to SCBV infection in comparison with the germplasm collections. The majority of the novel genotypes reported from the study was from germplasm collections indicating a higher percentage of nucleotidal variation in the viral genome infecting the germplasm. The germplasm clones have come from different regions of South East Asia, New Guinea, India, China and other regions [31] and they carry the varying virus population originated from the respective locations. Whereas, the hybrid varieties are evolved from Coimbatore, the major sugarcane breeding centre in India and with a limited virus population in the parental clones, mostly of hybrid varieties; the same virus population is expected in the new hybrid varieties through maternal transmission of the virus.

RT/RNase H region, a common taxonomic marker used for badna viruses, has been exploited to assess diversity among SCBV species. In our study, we found PCR amplification of 794 bp fragment of the virus in 48 and 57% of the samples from germplasm clones and *Saccharum* hybrid cultivars; respectively. Amplified fragments (~ 6070–6864 regions in RT/RNase H) were confirmed to be SCBV sequences with 80–99% similarity to RT/RNase H region of ORF3 polyprotein. Phylogenetic analysis with the obtained sequences revealed five novel genotypes viz. SCBV U-Y with < 92% nucleotidal similarity. The discrepancy within the nucleotide is distinct enough to establish that subclade as novel SCBV genotypes. SCBV-W forms a separate subclade in the *Badnavirus* group 1, while the other four new genotypes segregated in different sub clades in the *Badnavirus* group 3. Findings of the present study are evident with pairwise sequence identity analysis, which pointed out a lower intra-genotype identity when compared to already existing genotypes-SCBV A-T [1, 10, 19, 32]. SCBV-W would be assigned as a new species due to its genetic divergence > 22% at the RT/RNase H region. However, it necessitates the analysis of more nucleotidal sequences from the demarcation region or complete sequence analysis of SCBV-W: CBJ 46 isolate. All the 104 isolates from the study were segregated into the following genotypes/subgroups-

SCBV-E, -G, -H, -I, -J, -L, -Q, -R, -S and -T apart from the new SCBV U-Y genotypes. The distribution of 15 SCBV genotypes within 104 isolates revealed the genetic diversity within SCBV species. To date, 19 genotypes are found from India, excluding SCBV-A, -B, -C, -D, -N, -O, and -P. However, further analysis of SCBV population from unrepresented regions and germplasm clones may throw more light on the viral genomic diversity.

Recombination in viruses is a pervasive process led to the genetic diversity in most viruses. Since these pararetroviruses are well-known for their diversity, efforts were made to ascertain recombination events, which can be crucial for their distinct genomic behaviour. Potent recombination events were observed in SCBV isolates, especially from the proposed new genotypes - SCBV-U, -W and -X. Establishing 59 isolates, Subgroup U with potent recombinants and slight recombinants, points out their existence might have happened due to the genetic reassortment over time. SCBV-U, one of India's predominant genotype, also contributed mainly as a major or minor parent for the other genotypes. SCBV-L genotype proposed earlier by Rao *et al.* [21] as a new species is also reported from the current study. Seventeen isolates from the study segregated with SCBV-L, the second-highest reported genotype from Indian states. Isolates belonging to the Indian and Chinese domain shared a superior genetic similarity compared to the Guadeloupe / Australian origin isolates.

Even though the recombination phenomenon happened randomly or non-randomly, selective pressures must have acted against the breakpoints at definite position in the genome [17]. Negative values obtained in Tajima's D and Fu and Li neutrality tests indicate the existence of low-frequency polymorphism in the SCBV population. In contrast to the present study, Rao *et al.* [21] reported a positive D value of 2.68, an indication of balancing selection and deep subdivision in the SCBV population. The estimation of the ratio of non-synonymous (dN) to synonymous (dS) substitution of  $< 1$  revealed shreds of evidence of invasiveness. Purifying selection of SCBV population prevents the change of an amino acid residue, thereby favouring an excess of synonymous substitution over non-synonymous substitutions. The genomes from the same species vary in sequence as a result of different evolutionary processes. Codon usage bias is an effective tool for identifying patterns, hypothesizing that codon bias exists because of non-randomness in the mutational events. In other words, some codons can undergo more changes and therefore result in equilibrium frequencies of the codon. RSCU value obtained from the study ensured over expression of codon like AGA (Arginine) in all the SCBV genotypes and other hands, under-representation of particular codons. The frequency of a few codons is evident among all the 25 SCBV genotypes, constituting their diversity. In abbreviation, the current study successfully pointed out the possible reasons behind the genetic diversity found in SCBV species/genotypes with the aid of recombination, neutrality test and codon usage bias analysis.

During the first report of the virus in India it was found in few sugarcane germplasm clones [28], however, subsequently its widespread occurrence in various species clones of *Saccharum*, allied genera and hybrids clones of Indian and foreign origin [29, 31] maintained in the world sugarcane germplasm collections. Recent studies have clearly established that the virus infects the crop across 5.2 M Ha sugarcane growing areas in India and rampant occurrence of the disease has become a cause of concern

to sugarcane production [21, 30], apart from its impediment in germplasm exchange. At this scenario, there is a need to document the genotypic variation in SCBV, hence we have studied the genetic variation among the SCBV population with 104 isolates and it revealed existence of 15 genotypes of the virus, including the five new genotypes. Based on complete genome of the virus, earlier we have documented prevalence of five SCBV species, indicating enormous virus diversity in the country [10].

## Conclusion

Genome sequences surfaced as the primary and principal tool for the characterization of viruses during the last decades. The present study established genomic variation in SCBV isolates infecting a large number of *Saccharum* spp. clones and hybrids varieties in India and this brought out detailed information on variation in SCBV species in the country. With phylogenetic evidence and sequence demarcation tool, five novel SCBV genotypes: SCBV-U, -V, -W, -X and -Y were proposed in addition to 20 genotypes already established. Because of the genetic divergence with other existing isolates and repudiating sequence similarity to any of the whole genome sequencing available in NCBI data, SCBV-W showcased itself as a potential candidate for novel species. Considering the published data of SCBV isolates over the globe and analysis from the study suggest; Indian SCBV isolates as the most divergent population of SCBV in comparison with other regions. With the number of genotypes reported from Indian sugarcane materials chiefly from germplasm clones, there is a high probability of additional variants if a more number of clones are screened and analysed for genotypic variation in SCBV. Recombination patterns from the study will light on the probable heterogeneity and exchange of gene fragments between the parental sequences, leading to the evolution of new SCBV variants. There was no evidence of genetic recombination or gene assortment found between the SCBV A-D genotypes and SCBV E-Y isolates. However, this may have piloted the absence of former genotypes from the Indo-China region. Interestingly, Indian isolates showed considerable similarity to SCBMV (NC 008017) and SCBIMV (AJ277091) isolates from Morocco and Australia, respectively. Low polymorphism indicating a recent selective swap or purifying selection obtained from a functional part of the genome and a probable gene assortment/genetic drift that results in adventurous effect in SCBV genome pattern. In conclusion, our results emphasize the nucleotidal variation within SCBV species, thus aid in the molecular screening of commercial sugarcane material and germplasm clones in quarantine. The population genetic parameters and recombination signals can provide evolutionary details; thereby lend a hand in the parallel development of taxonomic studies. Perpetually increasing the deposit of genome sequences in the depository and identifying new genotypes necessitate the importance of phylogenetic subgroups/genotypes to be identified to the level of species/strains in the genus *Badnavirus*.

## Declarations

### Funding

The research work was supported by the Department of Science and Technology, Government of India New Delhi, under the grant SERB-EMR/2017/003285.

## Conflict of interests

The authors with this declare no conflict of interest regarding the publication of the article.

## Availability of data

All the 104 partial viral sequences determined in this study have been deposited in the GenBank database (Accession no: MW463323- MW463332, MW521165- MW521174, MW530428- MW530457, MW548458- MW548487, MW584705- MW584721, MW645068- MW645074).

## Ethics approval

The author declare that they have no conflict of interest.

## Consent to participate

All the authors have read the manuscript and consented this publication.

## Consent for publication

Informed consent was obtained from all individual participants included in the study

## References

1. Ahmad K, Sun S-R, Chen J-L, Huang M-T, Fu H-Y, Gao S-J (2019) Presence of diverse sugarcane bacilliform viruses infecting sugarcane in China revealed by pairwise sequence comparisons and phylogenetic analysis. *The Plant Pathol J* 35:41–50
2. Ashraf MA, Feng X, Hu X, Ashraf F (2020) An Algorithmic framework for genome-wide identification of sugarcane (*Saccharum officinarum* L.)-encoded microRNA targets against SCBV. LSSZ bioRxiv, pp 1–38
3. Balan S, Nithya K, Cherian KA, Viswanathan R (2021) True seed transmission of *Sugarcane bacilliform virus* (SCBV) in sugarcane. *Sugar Tech*. 10.1007/s12355-021-01031-0
4. Borah BK, Sharma S, Kant R, Johnson AA, Saigopal DVR, Dasgupta I (2013) Bacilliform DNA-containing plant viruses in the tropics: commonalities within a genetically diverse group. *Mol Plant Pathol* 14:759–771
5. Bouhida M, Lockhart BEL, Olszewski NE (1993) An analysis of the complete sequence of a sugarcane bacilliform virus genome infectious to banana and rice. *J Gen Virol* 74:15–22
6. Da Silva JM, Jobim LJ, Ramos Sobrinho R, Lima JS, Assunção IP, Cruz MM, Lima GS (2015) Incidence and species diversity of badnaviruses infecting sugarcane from a germplasm collection in Brazil. *Trop Plant Pathol* 40:212–217
7. Fernandez E, Ferdinand R, Filloux D (2020) Improvements in virus detection at CIRAD's sugarcane quarantine using both viral metagenomics and PCR-based approaches. *Sugar Tech* 22:498–503

8. Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* 133:693–709
9. Geering ADW, Hull R (2012) Family *Caulimoviridae*. In A. M. Q. King, M. J. Adams, E. B. Carestens, E. J. Lefkowitz (Eds.), *Virus taxonomy classification and nomenclature of viruses ninth report of the international committee on taxonomy of viruses*. San Diego: Elsevier pp:424–443
10. Karuppaiah R, Viswanathan R, Ganesh Kumar V (2013) Genetic diversity of *Sugarcane bacilliform virus* isolates infecting *Saccharum* spp. in India. *Virus Genes* 46:505–516
11. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol* 35:1547–1549
12. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452
13. Lockhart BEL, Autrey LJC (1991) Mealybug transmission of sugarcane bacilliform and sugarcane clostero-like viruses. In: 3rd ISSCT Sugar Cane Pathology Workshop pp. 17
14. Lockhart BEL, Autrey LJC (2000) Sugarcane bacilliform virus. In Rott P, Bailey RA, Comstock JC, Croft BJ, Sauntally AS (Eds.), *A guide to sugarcane diseases CIRAD and ISSCT* pp:266–272
15. Lockhart BEL, Irey MS, Comstock JC (1996) Sugarcane bacilliform virus, Sugarcane mild mosaic virus, and Sugarcane yellow leaf syndrome. In: Croft BJ, Piggitt CT, Wallis ES, Hogarth DM (eds) *Sugarcane Germplasm Conservation and Exchange*. Australian Centre for International Agricultural Research, Canberra, pp 108–112
16. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol* 1:vev003
17. Martin DP, Rybicki EP (2002) Investigation of Maize streak virus pathogenicity determinants using chimaeric genomes. *Virology* 300:180–188
18. Muhire BM, Varsani A, Martin DP (2014) SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS ONE* 9(9):e108277
19. Muller E, Dupuy V, Blondin L, Bauffe F, Daugrois JH, Nathalie L, Iskra-Caruana ML (2011) High molecular variability of sugarcane bacilliform viruses in Guadeloupe implying the existence of at least three new species. *Virus Res* 160:414–419
20. Pond SL, Frost SD (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21(10):2531–2533
21. Rao GP, Sharma SK, Singh D, Arya M, Singh P, Baranwal VK (2014) Genetically diverse variants of sugarcane bacilliform virus infecting sugarcane in India and evidence of a novel recombinant *Badnavirus* variant. *J Phytopathol* 162:779–787
22. Rodriguez-Lema E, Rodriques D, Fernandez E, Acevedo R, Lopez D (1985) Reporte de un nuevo virus de la cana de azucar *Ciencias de la Agricultura* 23:130
23. Sharp PM, Li WH (1987) The codon adaptation index—a measure of directional synonymous codon usage bias and its potential applications. *Nucleic Acids Res* 15(3):1281–1295



24. Sun SR, Damaj MB, Alabi OJ, Wu XB, Mirkov TE, Fu HY, Chen RK, Gao SJ (2016) Molecular characterization of two divergent variants of sugarcane bacilliform viruses infecting sugarcane in China. *Eur J Plant Pathol* 145:375–384
25. Tajima F (1989) The effect of change in population size on DNA polymorphism. *Genetics* 123(3):597–601
26. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10:512–526
27. Viswanathan R (1994) Detection of sugarcane virus and MLO diseases. Sugarcane Breeding Institute, Coimbatore, Annual Report for 1993–1994:56
28. Viswanathan R, Alexander KC, Garg ID (1996) Detection of sugarcane bacilliform virus in sugarcane germplasm. *Acta Virol* 40:5–8
29. Viswanathan R, Balamuralikrishnan M, Premachandran MN, Tripathi BK (1999) Sugarcane bacilliform virus: Symptoms, detection and distribution in the world germplasm collection at Cannanore. *Proc Intern Soc Sugar Cane Technol* 23:347–354
30. Viswanathan R, Karuppaiah R, Bagyalakshmi K, Sanju B, Kaverinathan K (2019) Emergence of leaf fleck caused by Sugarcane bacilliform virus in sugarcane as a serious disease under field conditions in India. *Intern Sugar J* 121:146–153
31. Viswanathan R, Premachandran MN (1998) Occurrence and distribution of sugarcane bacilliform virus in the germplasm collection in India. *Sugar Cane* 6:9–18
32. Wu XB, Alabi OJ, Damaj MB, Sun SR, Mirkov TE, Fu HY, Chen RK, Gao SJ (2016) Prevalence and RT/RNase H genealogy of sugarcane bacilliform virus isolates from China. *J Phytopathol* 164:595–607

## Figures

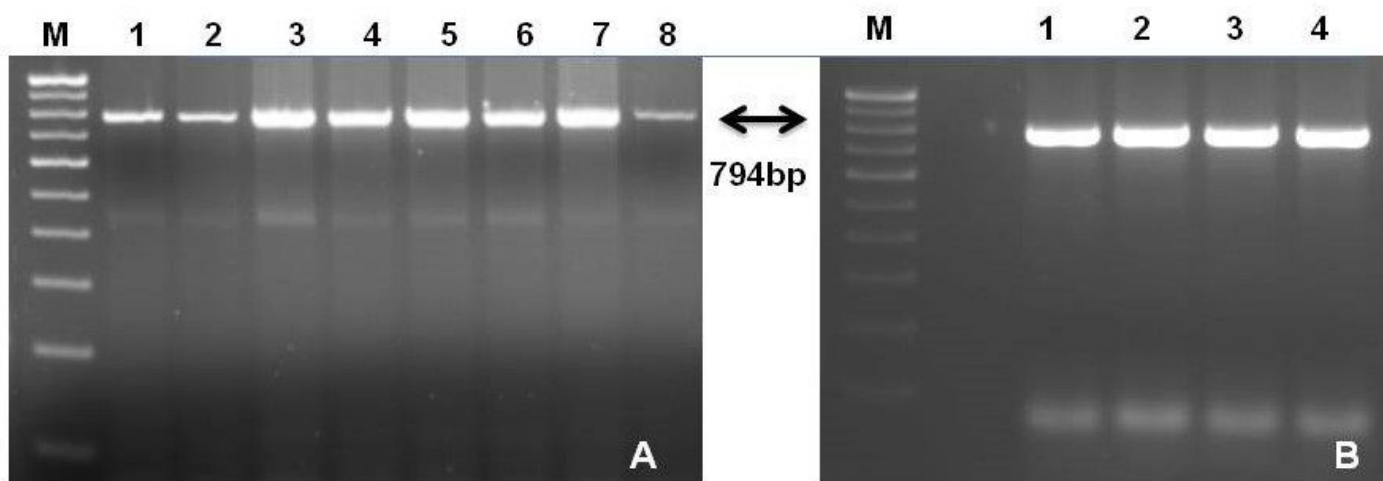
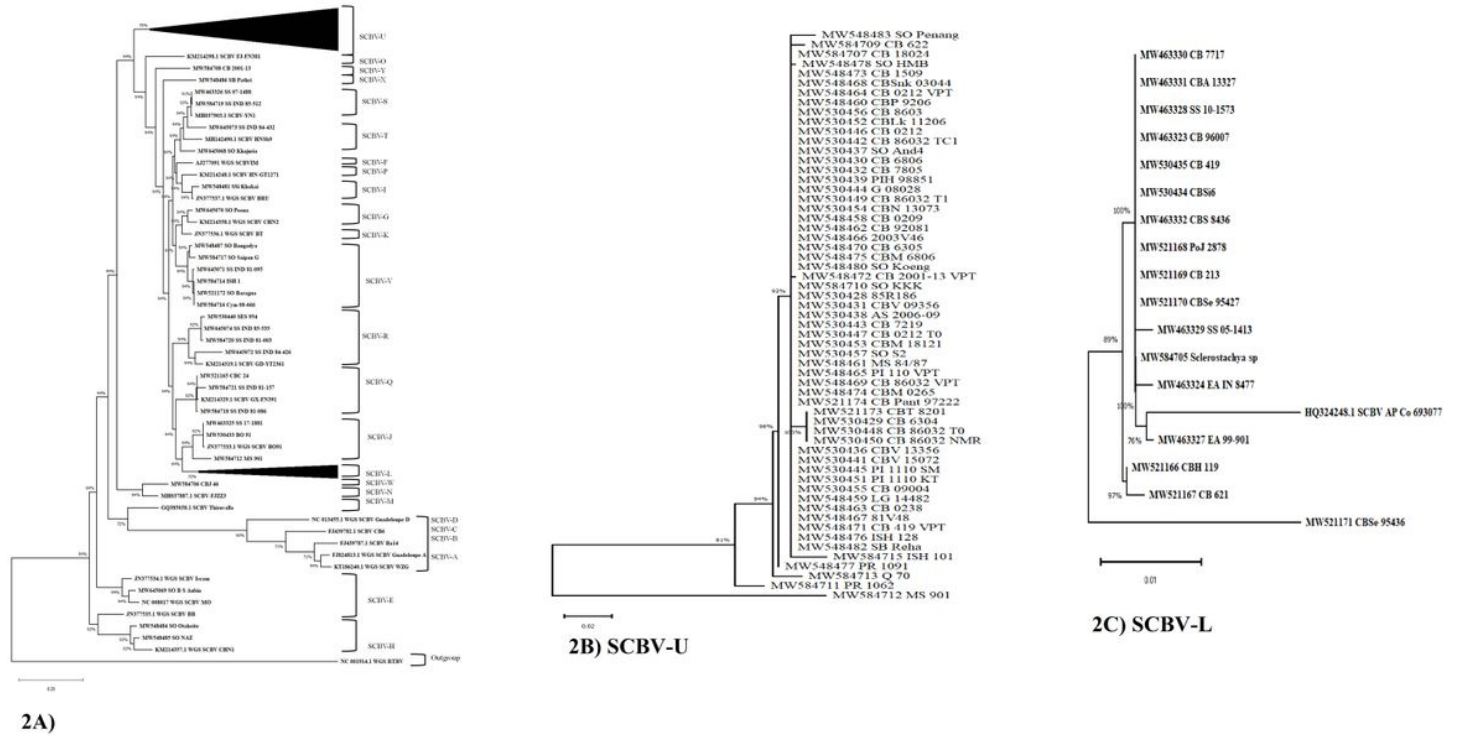


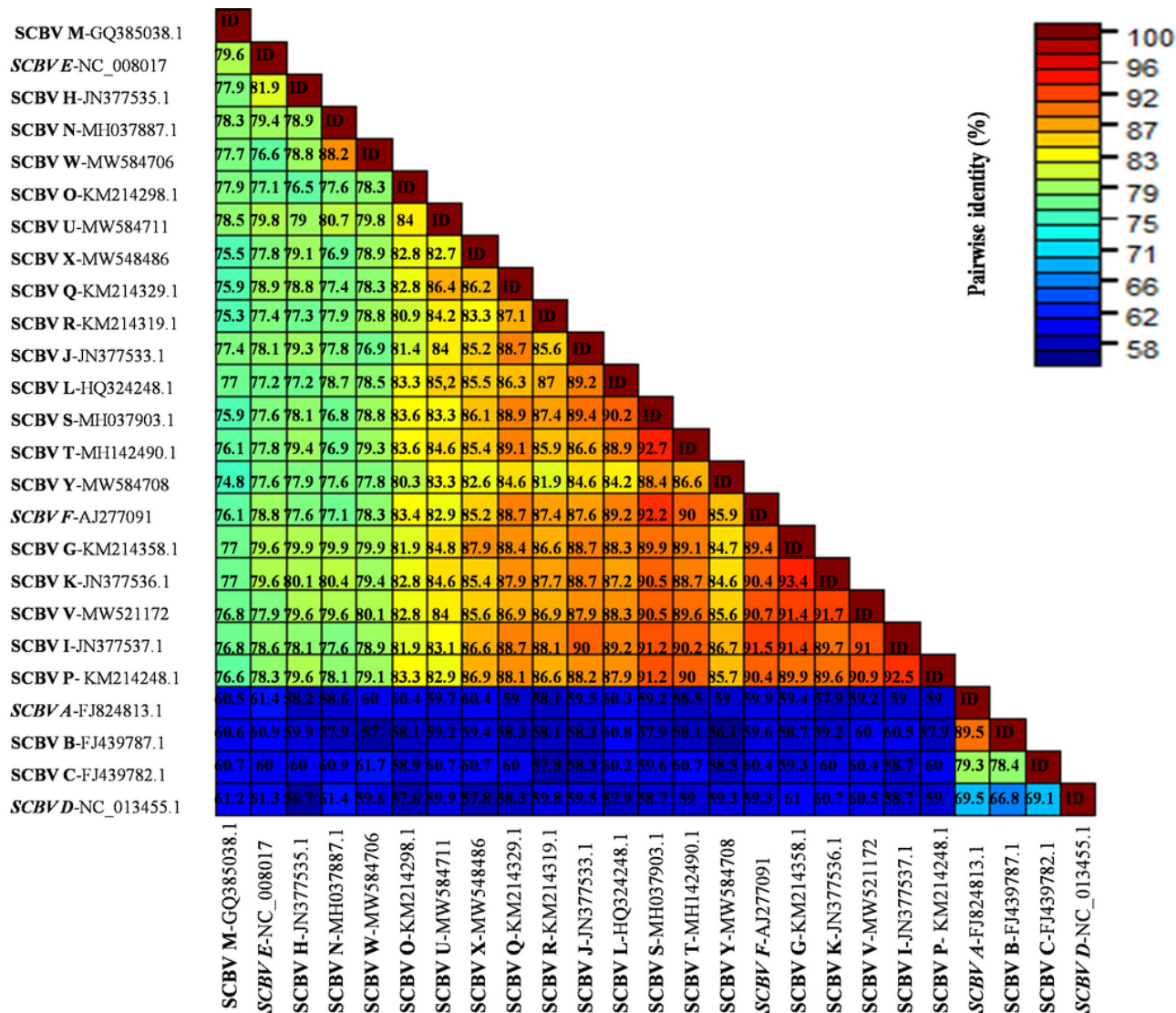
Figure 1

Agarose gel (1.5%) showing PCR amplified fragments ~794bp specific to SCBV from sugarcane germplasm and hybrid varieties A) Germplasm samples : 1- SS Khakai, 2-SO NAZ, 3-SO Baragua, 4- SS Koeng, 5-SO Bangadya, 6-SB Pathri, 7-MS 901, 8- SES 954. B) Samples from Saccharum hybrids: 1- CB 96007, 2-B091, 3- CBC 24, 4- PoJ 2878



**Figure 2**

Phylogenetic analysis of Sugarcane bacilliform virus isolates based on RT/RNase H region of ORF 3 polyprotein-603nt. A) The phylogeny of 128 isolates constructed using the Tamura-Nei model. The bootstrap consensus tree was inferred from 1000 replicates. B) Details of 59 samples, which clustered together to form SCBV-U genotype. C) Details of 17 isolates from the study, clustered together with SCBV AP Co 693077 from SCBV-L genotype.



**Figure 3**

Colour coded representation of pairwise percentage nucleotidal similarity among RT/RNase H region of Sugarcane bacilliform virus genotypes emerged from the study with genotypes elsewhere in the world. Colour scales in accordance with the percentage are showed. Acronyms are as follows: SCBV A- SCBGAV, SCBV B- SCBV BA14, SCBV C- SCBV CB6, SCBV D- SCBGDV, SCBV E - SCBMOV, SCBV F- SCBIMV, SCBV G- SCBV-CHN2, SCBV H- SCBBBV, SCBV I- SCBBRV, SCBV J- SCBBOV, SCBV K-SCBV BT, SCBV L- SCBV Co 693077, SCBV M- SCBV Thiruvalla, SCBV N- SCBV FJZZ3, SCBV O- SCBV FJ-FN381, SCBV P- SCBV HN-GT1271, SCBV Q- SCBV GX- FN 391, SCBV R- SCBV GD-YT2361, SCBV S- SCBV YNI, SCBV T- SCBV HNSb9, SCBV U- SCBV PR 1062, SCBV V- SCBV SO Baragua, SCBV W- SCBV CBJ46, SCBV X- SCBV SB Pathri, SCBV Y- SCBV CB2001-13. ICTV approved species are given in italics. GenBank accession number of each genotype showed in the matrix.

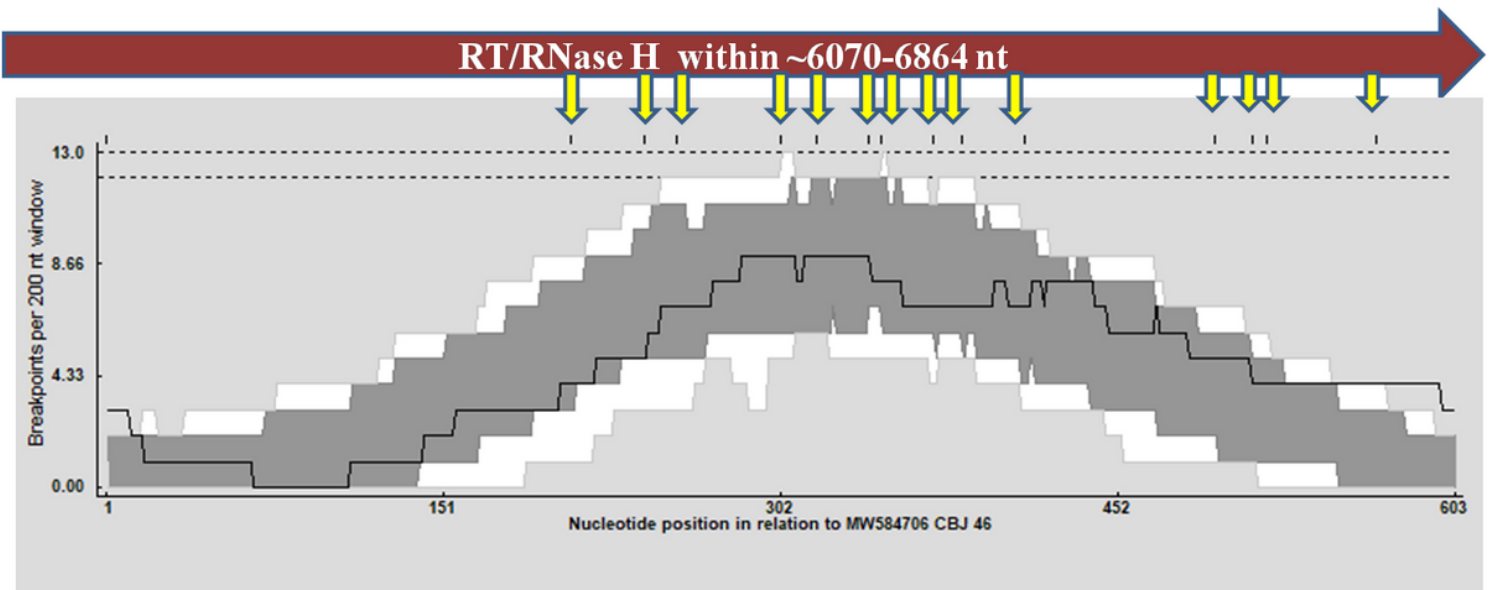
```

          10          20          30          40          50          60          70          80
SCBV A FJ824813.1 VQQHKEHLKK FMEICERNGL VLSPTKMKIG TRQVDFLGAT IGNSKIKLQP HIIQKIIDIK DEELKEVKGL RKWLGILNYA
SCBV B FJ439787.1 .....I .....I .....EM. ....
SCBV C FJ439782.1 .....AET. ....FR. .R. ....
SCBV D NC_013455.1 P...VQ.... .....K.... VS..... Q...R... ..K..A.FE ..K...T... .S.....
SCBV E NC_008017 EE...A...W MLQ...K... I...S.Y... VKK...S... ..DNQLAV.E ..K...AEFD ..K...TKE.. KS..AT....
SCBV F AJ277091 EEK.A...W MLQ...KK... I...S.Y... VKR...S... ..DNQLAV.E ..VS..A.FD E.R...TKE.. KS..AT....
SCBV G KM214358.1 EEE.A...WV MLQ...KK... I...S.Y... VKR...S... ..DNQLTV.D ..VS..A.FD ..R...TKE.. KS..AT....
SCBV H JN377535.1 EEE.A...WA MLQ...KK... I...S.Y... VKR...S... ..DNQLAV.D ..S..A.YD ..K...TKE.. KS..AT....
SCBV I JN377537.1 EEK.A...WI MLQ...KK... I...S.Y... VKR...S... ..DNQLAV.E ..VS..A.FD E.R...TKE.. KS..AT....
SCBV J JN377533.1 EDE.A...WI MLQ...KK... I...S.Y... VKR...S... ..DNQLAV.E ..VS..T.FD E.R...TKE.. KS..AT....
SCBV K JN377536.1 EEE.A...WI MLQ...QK... I...S.Y... VKR...S... ..DNQLAV.D ..VS..A.FD ..K...TKE.. KS..AT....
SCBV L HQ324248.1 EEE.A...WI MLQ...KK... I...S.Y... VKR...S... ..DNQLAV.E ..VS..A.FD E.R.RTKE.. KS..AT....
SCBV M GQ385038.1 EEE.V...W MLQ...KE... I...S.Y... VKK...S... ..DNMLQV.E ..T..A.FS ..K...TKE.. KS..AT....
SCBV N MH037887.1 EEE.A...WE MLK...KK... I...S.Y... VKR...S... ..DNQLAV.E ..S..A.FS ..K...TKE.. KS..AT....
SCBV O KM214298.1 EEE.A...W MLQ...KD... I...S.Y... VKR...S... ..DNQLSV.E ..VS..T.FD E.K...TKE.. KS..AT....
SCBV P KM214248.1 EEK.A...WI MLQ...KK... I...S.Y... VKR...S... ..DNQLAV.E ..VS..A.FD E.R...TKE.. KS..AT....
SCBV Q KM214329.1 EED.A...WT MLQ...KK... I...S.Y... VKR...S... ..DNQLAV.E ..VS..AEFD ..K...TKE.. KS..AT....
SCBV R KM214319.1 EEE.A...WI MLQ...KK... I...S.Y... VKR...S... ..DNQLAV.E ..VS..A.FD E.R...TKE.. KS..AT....
SCBV S MH037903.1 EEE.A...WV MLQ...KK... I...S.Y... VKR...S... ..DNQLAV.E ..VS..A.FD E.R...TKE.. KS..AT....
SCBV T MH142490.1 EEE.A...W MLQ...KK... I...S.Y... VKR...S... ..ENQLAV.E ..VS..A.FD ..K...TKE.. KSR.AT....
SCBV U MW584711 EED.A...WN MLQ...KK... I...S.Y... VKR...S... ..DNQLAV.E ..VS..A.FD E.K...TKE.. KS..AT....
SCBV V MW521172 EEK.A...WI MLQ...KK... I...S.Y... VKR...S... ..DNQLTV.E ..S..A.FD ..R...TKE.. KS..AT....
SCBV W MW584706 EEE.A...WE MLK...KK... I...S.Y... VKR...S... ..DNQLAV.E ..S..A.FS ..K...TKE.. KS..AT....
SCBV X MW548486 EEE.A...W MLQ...K... I...S.Y... VKR...S... ..DNQLV.E ..VS..A.FD E.R...TKE.. KS..AT....
SCBV Y MW584708 EEE.A...W MLQ...KK... I...S.Y... VKR...S... ..ENQLAV.E ..VS..A.FD ..K...TKE.. KI..AT....
          90          100          110          120          130          140          150
SCBV A FJ824813.1 RSYIPKLGKI LGPLYSKTSP NGERRMNSQD WKIVKEVKEI VANLPDLELP PEKAIMIET DGCMDDGWWGV CK
SCBV B FJ439787.1 .....D...M... ..E...
SCBV C FJ439782.1 .....N... ..N... ..R... .V..DKI.NL .K...E... ..NN.M... ..E...AI ..
SCBV D NC_013455.1 .N...N...T .....V... T...K...H...VLINK..RQ .KE.E.M... ..VE.CIVL... ..E...
SCBV E NC_008017 .NH.KDM..L .....P...E K...L..E... ..LINRI.TM .RT..N.TI... ..D.YI... ..A.AT...A...
SCBV F AJ277091 .NH.KDM..L .....P...E K.AKGL..E... ..LISRI.TM .RS..S.TI... ..D.YI... ..A.AT...A...
SCBV G KM214358.1 .NH.KDM..L .....P...E K.AKGL..E... ..LISRI.TM .R...K.TI... ..D.YI... ..A.AT...A...
SCBV H JN377535.1 .NH.KDM..L .....P...E K...L..E... ..LISRI.TM .RS..K.TIS... ..D.YI... ..A.AT...A...S...
SCBV I JN377537.1 .NH.KDM..L .....P...E K.AKGL..E... ..LISRI.TM .R...S.TI... ..DD.YI... ..A.AT...A...
SCBV J JN377533.1 .NH.KDM..L .....P...E K.AKGL..E... ..LISRI.TM .RS..N.TI... ..D.YI... ..A.AT...A...
SCBV K JN377536.1 .NH.KDM..L .....P...E K.AKGL..E... ..LISRI.TM .R...K.TI... ..D.YI... ..A.AT...A...
SCBV L HQ324248.1 .NH.KDM..L .....P...E K.AKGL..E... ..LISRI.TM .R...N.TI... ..D.YI... ..A.AT...A...
SCBV M GQ385038.1 .NH.KDM..L .....P...E K...L..AE... ..LIYKI.VL .K...K.SI... ..D.YI... ..A.AT...A...
SCBV N MH037887.1 .NH.KDM..L .....P...K...L..E... ..LIGRI.TM .R...N.TI... ..DD.YI... ..A.AT...A...
SCBV O KM214298.1 .NH.KDM..L .....P...E K.A..GL..E... ..LISRI.TM .RS..N.SI... ..DD.YI... ..V.AT...A...
SCBV P KM214248.1 .NH.KDM..M .....P...E K.AKGL..E... ..LISRI.TM .R...K.TI... ..DD.YI... ..A.AT...A...
SCBV Q KM214329.1 .NH.KDM..M .....P...E ..AKGL..E... ..LISRI.TM .RS..N.TI... ..D.YI... ..A.AT...AI...
SCBV R KM214319.1 .NH.KDM..L .....P...E K.AKGL..E... ..LISRI.TM .RS..N.TI... ..D.YV... ..A.AT...A...R...
SCBV S MH037903.1 .NH.KDM..L .....P...E K.AKGL..E... ..LISRI.TM .R...N.TI... ..D.YI... ..A.AT...A...
SCBV T MH142490.1 .NH.KDM..L .....P...E K.AKGL..E... ..LISRI.TM .R...N.TI... ..D.YI... ..A.AT...A...
SCBV U MW584711 .NH.KDM... ..P...E K...F..E... ..LINRI.TM .RT..N.TI... ..D.YL... ..A.AT...A...
SCBV V MW521172 .NH.KDM..L .....P...E K.AKGL..E... ..LISRI.TM .IR...S.TI... ..DD.YI... ..A.AT...A...
SCBV W MW584706 .NH.KDM..L .....P...E K...L..E... ..LINRI.TM .RS..N.TI... ..DD.YI... ..V.AT...A...
SCBV X MW548486 .NH.KDM..L .....P...E K.AKGL..E... ..LISRI.TM .RS..K.TI... ..D.YI... ..A.AT...A...
SCBV Y MW584708 .NH.KDM..L .....P...E K.KKDL..E... ..LISRI.TM .RS..N.TI... ..DD.YI... ..ASATR..A...

```

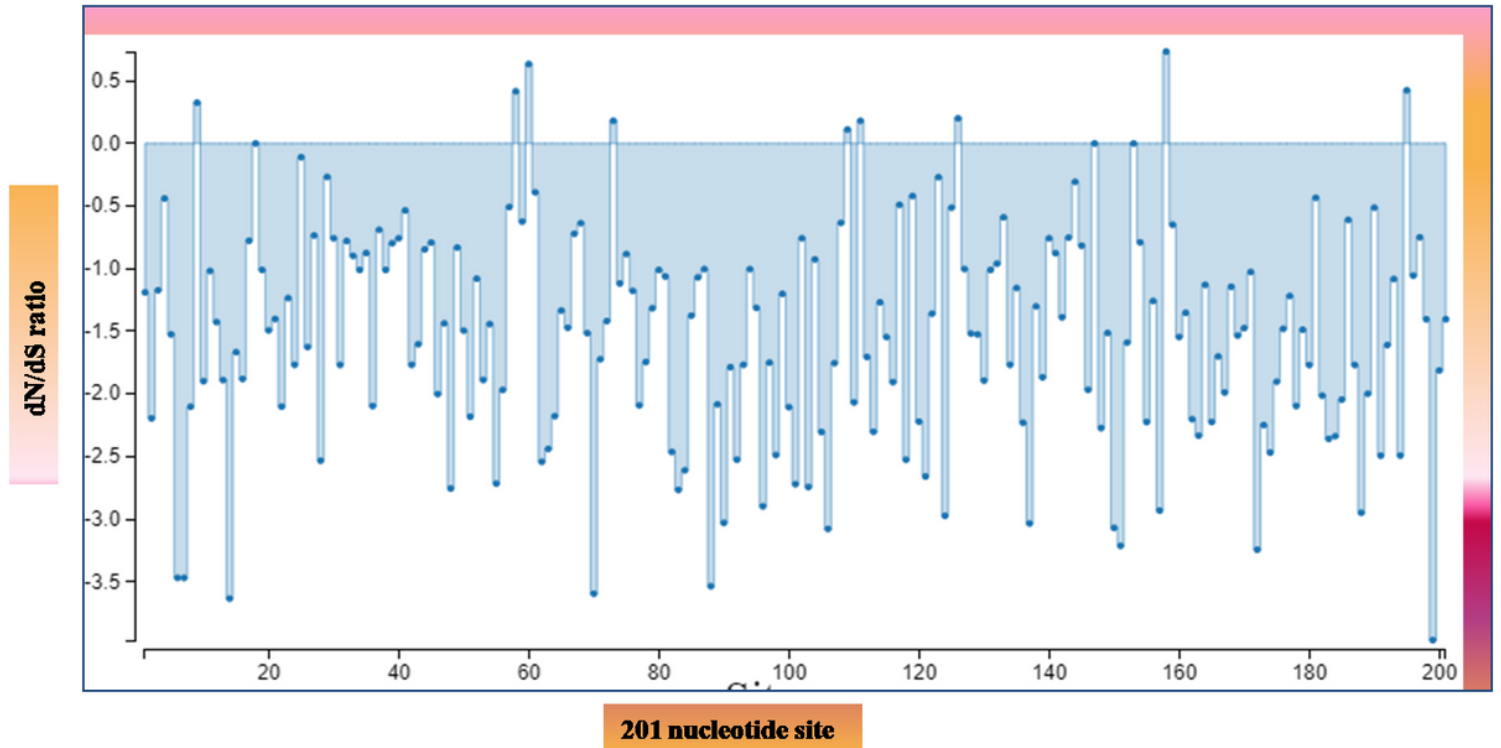
Figure 4

Amino acid sequences alignment with variations in the RT/RNase H-region of SCBV genotypes; SCBVA – Y with Genbank accession numbers.



**Figure 5**

Recombination patterns amongst Sugarcane bacilliform virus using RDP4 detected in the region-ORF3: RT/RNase H. Vertical arrows with yellow colour represent unique breakpoints with in the 124 SCBV sequences in the study. For determining hot and cold spots of recombination, a 200 nt window was moved along CBJ46 isolate. Light and dark grey areas symbolize local 99% and 95% breakpoint clustering, respectively.



**Figure 6**



Selection pressure on the coding region of SCBV genotypes by SLAC method (Single-Likelihood Ancestor Counting). The number of synonymous and non-synonymous alterations which have occurred at each site within a total of 201 sites; the ratio of dN/dS ( $\omega = dN/dS$ ) are depicted in the graph, where dN- non synonymous substitution and dS- synonymous substitution.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.docx](#)