

# Evidence for Recombination in the Emerged Phylogroup SW6 of *Prunus Necrotic Ringspot Virus* and Estimate of Selection Pressure Acting on Capsid Protein of Affiliated Members

Moncef Boulila<sup>1\*</sup>

<sup>1</sup>Institut de l'Olivier B.P. 14; 4061 Sousse Ibn-Khaldoun, Tunisia  
Email: [boulila.moncef@yahoo.fr](mailto:boulila.moncef@yahoo.fr)

**Abstract.** Recombination plays a key role in virus evolution. Except a limited number of works carried out so far on recombination in *Prunus necrotic ringspot virus* (PNRSV) genome, knowledge is still poor. Numerous phylogenetic studies particularly those based on capsid (CP) and movement protein (MP) of PNRSV showed that all accessions described so far split into three major phylogroups, i.e., PV96, PV32, and PE5. With the recent characterization of Chilean and Indian isolates which were phylogenetically closely related to the oldest member, i.e., SW6 (AF013287) from USA, a new phylogroup baptized SW6, emerged since it was outside of the evolutionary behavior of the classical PV96, PV32, and PE5 clusters. In order to understand how it evolved, six computerized methods were used to detect potential recombination in the CP sequences of nine components forming SW6 phylogroup. It was clearly demonstrated that various members underwent recombination. Additionally, networked relationships among them proved that numerous incompatibilities occurred in these sequences illustrated by the presence of boxes in the network implying therefore the possibility of recombination. According to inferred network, several sequences contained conflicting signals as illustrated by various splits and edges with special reference to Chilean (EF565255, EF565256), polish (AF332614) and Indian (AM494934) isolates. It is worth noting that the phylogenetic network provided a different clustering compared to the bifurcating tree. In fact, three distinct groups were delineated with SplitsTree4 software instead of a globally homogenous ensemble generated by MEGA7 algorithm. Furthermore, this study pointed out that members of SW6 phylogroup were the only recombinants among 205 tested accessions confirming that SW6 should be considered as the fourth phylogenetic group of PNRSV. On the other hand, CP was predominantly under purifying selection. However, positive selection was evidenced particularly in the C-terminal region of CP comprising part of dimerization region (DR). Such adaptive selection was exerted on a CP segment of AM494934 accession exchanged by recombination.

**Keywords:** Recombination, network, bifurcating tree, selection, bioinformatics, PNRSV

## 1 Introduction

Populations of RNA viruses fundamentally consist of mutant clouds rather than genomes constituted by constant nucleotide sequence. These mutants are the source of virus adaptability since they form continuously changing repositories of viral variants. Virus evolution is decisively influenced by high mutation rate, molecular recombination and reassortment (for those having a divided genome). Recombination among plant RNA viruses plays a major evolutionary role in contributing and maintaining genetic diversity in virus population by creating mosaicism. Such opportunities can help viruses to overcome selection pressure and to adapt to new environments and hosts. Ignoring recombination may distort the analysis of genetic data and the conclusions derived from it. In segmented genomes as well as in unsegmented genomes, recombination takes place through template switching (or copy-choice model). The RNA-dependant RNA polymerase jumps from one RNA template to the other during replication. Furthermore, mutants are the target of selection. Selective pressure is any phenomena that alter the behavior and fitness of any living organism under the influence of biotic (living organisms within the same ecosystem that interact with the affected organism) or abiotic (light, wind, soil) factors. The selection pressure implies the opposition to the continued survival of such organism. Positive selection is the process by which advantageous mutations are fixed in the genes or

genomes, and consequently, they are responsible for evolutionary innovations and species divergence. In contrast, negative selection tends to eliminate deleterious mutations. However, negative selection does not mean that elimination can occur systematically. As a matter of fact, subsets of genomes could be maintained at low frequencies. At molecular level, Selection can be estimated through comparison of synonymous ( $dS$ ) and non-synonymous ( $dN$ ) substitution rates. If  $\omega = dN/dS = 1$ , amino-acid substitutions may be largely neutral. But, positive selection could possibly cancel negative selection. If  $\omega > 0$ , i.e.,  $dN/dS > 0$ , selection has caused some amino-acid substitutions. These substitutions may be caused by genetic drift or purifying selection but not strong enough to overcome the effect of positive selection. If  $\omega < 0$ , i.e.,  $dN/dS < 0$ , purifying selection has definitely operated. However, some amino-acid substitutions may be caused by positive selection, just not enough to overcome the effects of negative selection.

*Prunus necrotic ringspot virus* (PNRSV) is a positive-sense RNA plant virus with a tripartite genome that belongs to the genus *Ilarvirus* [1] which has the same genomic organization, encoding functionally similar translation products, as members of the genera *Bromovirus*, *Cucumovirus*, *Anulavirus*, *Oleavirus* and *Alfamovirus*, all of which belong to the family *Bromoviridae* [2]. RNA 1 and RNA 2 are monocistronic and encode nonstructural proteins (P1 and P2) involved in RNA replication [3]. RNA 3 is bicistronic and encodes two polypeptides, a polypeptide with homology to proteins required for cell-to-cell movement of plant viruses (ORF 3a) and the coat protein (ORF 3b) [4-7]. The movement protein (MP or P3) is translated from RNA3, while the coat protein (CP) synthesis occurs *via* a subgenomic monocistronic mRNA (RNA 4), which is collinear with the 3' end of RNA 3 and is encapsidated. Currently, the majority of reports on phylogenetic analyses based on coat protein-encoding gene mentioned that PNRSV clustered into three groups named PV32, PV96, and PE5 [8-11]. Despite the presence of the three phylogroups, there is no obvious correlation between clustering and geographic origin, symptomatology or host species. Nevertheless, most of isolates from the phylogroups PV96 and PV32 showed latent/mild or chlorotic/necrotic symptoms, respectively [12, 13]. Besides, Glasa *et al.* [14] proposed another phylogroup named CH30, but their results could be biased since they were based only on partially sequenced CP gene. On the other hand, Boulila [15] and Boulila *et al.* [16] revealed the presence of a supplementary phylogroup baptized SW6 which encompasses accessions from Chile, India, and Poland. Based on CP sequence analyses, this work aimed at characterizing this group at the molecular level by emphasizing its uniqueness as recombinant group in PNRSV and highlighting selective constraints exerted on capsid protein of members composing this phylogroup.

## 2 Material and Methods

The study sequences comprised nine PNRSV isolates whose CP sequences were extracted from GenBank (Table 1).

### 2.1 Alignment of Sequences and Construction of Phylogenetic Trees

Reconstruction of phylogenetic trees were made firstly by analyzing a total of 205 PNRSV coat protein sequences of all accessions deposited so far in the international databases in order to delimit the different various phylogroups. Secondly, phylogeny was inferred for all members composing SW6 cluster which was the subject of this study. As such, the nucleotide sequences of all 205 CP sequences were aligned using CLUSTAL Omega, CLUSTALX 2.1 [17] and Multalin [18] softwares with default setting. For the first case, the phylogenetic relationships among PNRSV isolates were determined with the Maximum Likelihood (ML) algorithm incorporated in the MEGA version 7 program [19]. Based on the evaluation of best fit substitution model executed in MEGA7, the ML tree was reconstructed under the assumption of substitution model K2 [20] coupled to a discrete Gamma distribution (+G) with five rate categories. The substitution model parameters estimated were (i) base frequencies:  $f(A) = f(T) = f(C) = f(G) = 0.250$ , (ii) substitution rates:  $r(AT) = r(AC) = r(TA) = r(TG) = r(CA) = r(CG) = r(GT) = r(GC) = 0.031$ ;  $r(AG) = r(TC) = r(CT) = r(GA) = 0.188$ ; and (iii) transition/transversion ratios:  $R = 2.98$ . The Bayesian Information Criterion value (BIC = 15403.079) with K2+G (+G = 0.40) model was the lowest among the 24 models tested. Bootstrap analyses were performed with 1,000 replicates to assess the robustness of the tree branches. Whereas for the second case, a phylogenetic tree was

generated from nucleotide alignment of nine members composing the SW6 group by means of MEGA7 software under the assumption of K2 as substitution model with uniformity of evolutionary rates among sites. Robustness of inferred evolutionary relationships was assessed by 1,000 bootstrap replicates. The parameters used were as following: (i) base frequencies:  $f(A) = f(T) = f(C) = f(G) = 0.250$ , (ii) substitution rates:  $r(AT) = r(AC) = r(TA) = r(TG) = r(CA) = r(CG) = r(GT) = r(GC) = 0.058$ ;  $r(AG) = r(TC) = r(CT) = r(GA) = 0.133$ ; and (iii) transition/transversion ratios:  $R = 1.14$ . The Bayesian Information Criterion value (BIC = 2999.183) was the lowest among the 24 models tested. Furthermore, in analyzing phylogenetic network, split neighborNet method, based on NJ distance method, and uncorrected P-distances algorithm implemented in SplitsTree4 program [21], were used. Distance in the evolutionary concept is the smallest amount of evolution that could have occurred. Each changed base pair is considered as one change; whereas, all unchanged base pairs are counted zero.

## 2.2 Recombination Analyses

Potential recombination events between diverged nucleotide sequences were explored with six programs: (i) SimPlot [22] by using the distance-based similarity method, allows to analyze a potential recombinant (or query sequence) against other sequences called *reference sequences* in order to identify (a) whether the query sequence is recombinant, (b) the position of potential breakpoints, and (c) putative minor and major parental; (ii) RDP 4.70 [23] incorporates several published recombination detection methods into a single suite of tools: RDP [24], GENECONV [25], BOOTSCAN [26], MAXCHI [27], CHIMAERA [28], SISCAN [29], and 3SEQ [30]. BOOTSCAN, RDP, and SISCAN are phylogeny-based methods; GENCONV, MAXCHI, CHIMAERA, and 3SEQ are substitution-based methods. In all cases, defaults parameters were used. Only events predicted by more than half of the methods are considered as significant; (iii) Bellerophon [31], a program for detecting chimeric sequences in multiple sequence datasets by an adaptation of partial treeing analysis based on distance-matrix calculation across breakpoints; (iv) DnaSP version 5, DNA Sequence Polymorphism [32]: a software package for the analysis of nucleotide polymorphism from aligned DNA sequence data. DnaSP can estimate several measures of DNA sequence variation within and between populations (in non-coding, synonymous or non-synonymous sites, or in various sorts of codon positions), as well as linkage disequilibrium, recombination, gene flow and gene conversion parameters based on population genetics; (v) SBP (Single Breakpoint) and (iv) GARD (Genetic Algorithm for Recombination Detection) [33, 34]. These two latter algorithms use a statistical approach to search recombination breakpoints from multiple-sequence alignments of homologous sequences. Potential breakpoints are identified by improvement of the small-sample corrected Akaike information criterion (AICc) [35] for phylogenetic trees constructed of individual recombinant fragments. Based on the outcome of the analysis, a level of support is assigned and expressed as a breakpoint placements score [33, 34]. Breakpoints identified by GARD were then assessed for significance using the KH test [36] of the HyPhy package [37].

## 2.3 Test of Neutrality and Evolution Assessment

DnaSP version 5.10.01 was used to estimate Tajima's  $D$  [37] and Fu and Li's  $D$  and  $F$  [39] statistical tests to examine the hypothesis of neutrality operating on the CP gene of nine sequences included in SW6 group. An estimate was also made of several population genetic parameters including nucleotide polymorphism ( $\pi$  estimated by the average number of nucleotide differences between two random sequences in a population), haplotype diversity ( $Hd$ , the frequency and number of haplotypes in a population), the statistic  $\theta$  from the number of segregation sites ( $S$ ) [40], the average rate of synonymous and non-synonymous substitutions,  $\Delta Hd$  (the variance of haplotype diversity),  $K$  (average of number of pairwise nucleotide differences). The distribution of  $dS$  and  $dN$  along the coding regions was analyzed using the SNAP program (<http://www.hiv.lanl.gov>; Korber [41]). Based on the results obtained by the statistical tests mentioned above and for a more extensive analysis, examination for selection was performed using codon-based Maximum Likelihood methods i.e., the Single-Likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), Random Effects Likelihood (REL) [42, 43], Internal Fixed Effects Likelihood (IFEL) [44], Mixed Effects Model of Episodic Selection (MEME) [45], Fast Unbiased Bayesian Approximation (FUBAR) [46], and the Partitioning Approach for Robust Inference of Selection (PARRIS) [47], models implemented at <http://www.datamonkey.org>, the web

server of HyPhy package [48]. To further investigate when and how selection pressure varied over the evolutionary history, GA-Branch (Genetic Algorithm-Branch) method [43], was applied. The GA-Branch program utilizes a genetic algorithm to test an extensive number of models of codon evolution based on small sample AIC score. This analysis permitted to classify each branch to a specific  $dN/dS$  rate class. Furthermore, PRIME (Property Informed Model of Evolution) determined which biochemical properties could drive substitutions at a given site; e.g. if a site is positively selected, then which properties are being selected for/against. The exchangeability function is a product of property-specific contributions. PRIME is a model which involves a parameter  $\alpha$  which represents the importance of property. Positive value ( $p < 0.05$ ) of  $\alpha$  cause the property to be conserved (purifying selection) whereas a negative value ( $p < 0.05$ ) means that the property tends to changing (positive selection). In case of  $\alpha = 0$ , selection is neutral with respect to that property. PRIME currently supports two predefined sets of five amino acid properties: the properties used by Conant et al. [49] ( $\alpha_1$ , chemical composition;  $\alpha_2$ , polarity;  $\alpha_3$ , volume;  $\alpha_4$ , iso-electric point;  $\alpha_5$ , hydrophathy), and Atchley et al. [50] ( $\alpha_1$ , polarity index;  $\alpha_2$ , secondary structure factor;  $\alpha_3$ , volume;  $\alpha_4$ , refractivity/heat capacity;  $\alpha_5$ , charge/iso-electric point). Recently, Kosakovsky Pond's team developed a new algorithm called BUSTED (Bayesian Unrestricted Test for Episodic Diversification) [51] (available in datamonkey website) aimed at detecting gene-wide evidence of positive selection that acts on a subset of branches in a phylogeny at a subset of sites within the gene by contrast to MEME method [46] which detects episodic diversifying selection at individual sites. In other words, to determine episodic positive selection in an entire gene, it is erroneous to combine a set of results output of MEME for drawing conclusions about a whole genome. Therefore, BUSTED is the best appropriate for that purpose.

### 3 Results

#### 3.1 Nucleotide Identity Comparison and Maximum Likelihood Estimate of Substitution Matrix and Transition/Transversion Bias

Nucleotide percent identity among nine PNRSV CP accession sequences varied from 88.90 to 100% (Table 2). Moreover, amino acid pairwise identity and similarity comparisons among these isolates indicated that they ranged from 87.11% to 100% and from 88.88% to 100%, respectively (Table 3). As a result, the most divergent accessions were AF332614 and AM494934 from one another. In contrast, the accessions AM408909 and AM419814 were identical. Substitution pattern and rates were estimated under the Kimura (1980) 2-parameter model. Rates of different transitional substitutions were 13.32 and those of transversional substitutions were 5.84. The estimated Transition/Transversion bias ( $R$ ) was 1.14. The nucleotide frequencies were A = 25.00%, T/U = 25.00%, C = 25.00%, and G = 25.00%. For estimating ML values, a tree topology was automatically computed. The maximum Log likelihood for this computation was -1429.908. The analysis involved nine nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 674 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.

#### 3.2 Phylogenetic Tree and Network Inference

To obtain information about evolutionary relationships among 205 PNRSV isolates based on CP sequences, a bifurcating phylogenetic tree was constructed using ML algorithm implemented in MEGA7. All analyzed sequences fell into three classical phylogroups (PV96, PV32, and PE5) along with a supplementary phylogroup named SW6 constituting the main subject of this study (Fig. 1). Additionally, they encompassed 121 (59%), 60 (29%), 15 (7%), and 9 (5%) accessions. As reported earlier, clustering did not reflect any correlation with specificity of host species, symptomatology, or geographic origin. Focusing on SW6, phylogenetic inference showed that all nine isolates constituted a quite homogenous cluster (Fig. 2). However, if the sequence alignment contains conflicting phylogenetic signals, a traditional phylogenetic tree becomes no longer sufficient to represent the evolutionary relationships among PNRSV isolates. Consequently, a phylogenetic network is much appropriate for that goal. Recombination is a reticulate event. Since recombination could be detected in different PNRSV isolates

belonging to SW6 group, SplitsTree4 program permitted to construct a split network which is a more general type of phylogenetic graph illustrating incompatibilities within and between data sets as indicated by the presence of boxes in the network (Fig. 3) implying therefore the possibility of recombination. According to inferred network, a few sequences contained conflicting signals as illustrated by various splits and edges. It is clear that the phylogenetic network provided a clustering quite different from the bifurcating phylogenetic tree. Indeed, the nine members segregated into three distinct clusters. Although, cluster I was composed by the accessions AF013287, EF565255, EF565256, AM408909, AM408910, AM419814 and AM920668, clusters II, and III comprised the accessions AF332614 and the accession AM494934, respectively. It is noteworthy that the length of the tree branches of the two latter accessions reflects the increasing of expected substitution sites within these sequences. The network is somewhat more complex to look at but seemingly, besides groupings, it showed clear similarities and differences between sequences.

### 3.3 Recombination Analyses

Recombination is a relatively common key factor of genetic diversity of RNA viruses and understanding this phenomenon is helpful in unraveling their evolution. In determining the occurrence of potential recombination events in multi-aligned nine sequences of SW6 group, six methods were used. According to DnaSP Version 5.10.01, at least two putative recombination events were detected. The recombination intervals had the positions 294-501, and 634-649. Similarly, Bellerophon detected two chimeras having the same position, i.e. 360. The two chimeras were the accessions EF565255 and EF565256. In addition, both chimeras generated from the contiguous segments of the parentals AM494934 and AF332614. The percent identities of parent sequences to chimera left and right of the breakpoint were 99.3% - 89.3% and 92.3% - 95.3% for AM494934 and AF332614 as parents 1 and 2, respectively. In corroborating these results, SimPlot method demonstrated that, indeed, EF565255 and EF565256 were two recombinants. The recombination breaking points for both accessions were 301 bp, 501 bp, and 521 bp (Fig. 4). As a matter of fact, there are more than two parents implicated in the genesis of the recombinants mentioned above. Besides AM494934 and AF332614 as two possible parentals, AM920668, AF013287, and AM419814 were supplementary donor and/or recipient of exchanged segments (Fig. 4). On the other hand, SimPlot algorithm pointed out that AM494934 accession was the progeny of the tentative parentals AM920668 and EF565255 lineages. Recombination signals occurred at the positions 161 bp and 381 bp (Fig. 5). It is worth mentioning that in spite of the fact that SimPlot procedure work well with highly diverged sequences, it suffers of inability to produce any measure of statistical confidence. In view of the fact that the evolution nature is stochastic, the distance of a sequence to other homologous sequence may fluctuate and interpretation could imply a somewhat uncertainty. Therefore, in order to overcome this shortcoming, the Recombination Detection Program (RDP v4.70) to identify potential recombination sites and possible minor and major parentals based on statistical support along with SBP and GARD algorithms which are incorporated in datamonkey web server, were used. Regarding RDP4.70 algorithm, a suite of five methods implemented in this software revealed that the progeny AM494934 was a recombinant having AM920668 as major parent. In contrast, the minor parent is still unknown (Fig 6). Nevertheless and according to SimPlot, the minor parent might be EF565255. The methods and ascribed  $p$ -values were as following: RDP ( $1.012 \cdot 10^{-8}$ ), GENCONV ( $5.421 \cdot 10^{-7}$ ), BOOTSCAN ( $3.035 \cdot 10^{-7}$ ), MAXCHI ( $2.864 \cdot 10^{-11}$ ), 3SEQ ( $3.108 \cdot 10^{-15}$ ). The recombination site was 608-673 bp. While SBP algorithm revealed a breakpoint in the position 613 supported by a corrected Akaike Information Criterion (cAIC) having the value 214.073, and a model averaged support of 100%, GARD placed a breakpoint at bp 179 position based on cAIC goodness of fit (cAIC = 2923.2).

### 3.4 Selection Estimate

In order to determine whether the number of segregating sites in the sequences departs from the neutral expectation, the software DnaSP version 5.10.01 was used. It permitted the calculation of Tajima's  $D$  as well as Fu and Li's  $D$  and  $F$  statistical tests to assess the neutrality and influence of demographic forces on the population which were as the following: Tajima's  $D = -1.75204$  (significant at  $p < 0.05$ ), Fu and Li's  $D = -1.94133$  and Fu and Li's  $F = -2.12993$  (significant at  $p < 0.05$ ) (Table 4). The calculation was based on the total number of mutations. The significantly negative values of Tajima's  $D$ , and Fu and

Li's  $D$  and  $F$  statistical tests for CP sequences discounted the neutral hypothesis suggesting the occurrence of purifying selection and demographic expansion of SW6 population of PNRSV based on CP analysis. Furthermore, the selection profiles of CP sequences were determined by submitting the sequence alignments to SNAP program where averages of all pairwise comparisons led to the conclusion that a purifying selection ( $dN < dS$ ) occurred. In fact the terms  $dN$  and  $dS$  were as following:  $dS = 0.0810$ ;  $dN = 0.0282$ ,  $dS/dN = 2.8723$  (Table 4). Thereafter, in the HyPhy package, available at the datamonkey server which implements various models of evolution, investigations site-by-site of the signature of selective pressure based on  $dN/dS$  ratio by applying the SLAC, FEL, IFEL, REL and FUBAR methods which incorporate non-synonymous as well as synonymous rate variation among codon sites, were conducted. Arising results indicated that a very limited number of codons were under purifying selection. Whereas SLAC and IFEL (0.1 significance level) (Tables 5a,b) each one detected one negatively selected codon, i.e., 98, and 219, respectively. FUBAR (posterior probability  $p \geq 0.9$ ) and FEL (0.1 significance level) models (Tables 5c,d) detected 3 and 15 negatively selected sites, respectively. It is clear that these two models detected in common three codons, i.e., 98, 167, 219. Additionally, the use of the PARRIS method resulted in no detection of positive selection at  $p < 0.1$  (Table 5e) in aligned sequences of CP gene as given by inferred distribution rates for the null (M1) and alternative models (M2) mentioned in Table 5e. Besides, REL ( $p = 0.02$ ) and MEME (0.1 significance level) models detected specifically 13 and 2 positively selected sites (Tables 5f,g), respectively. They revealed in common the codon 206. To acquire further insight into the lineage specific nature of the selective pressures acting on each branch of the phylogenetic tree, analyses using a genetic algorithm, namely GA-Branch, were performed. GA-Branch selected three classes with the support of 1517 models at 95% confidence set and a cAIC having the value 2862.83 ( $\Delta cAIC = 0.716515$ ). A total of 13 branches were enumerated (Fig. 7) and segregated into three parts. Whereas the first class was constituted by 5 branches, the second class was formed by 3 branches, and the third class comprised 5 branches whose  $dN/dS$  ratios were 1.21513, 1.0000, and 0.096451, respectively. In contrast, using PRIME model, the exchangeability function did not produce any property-specific contribution. In other words, no physico-chemical property of amino acids was able to drive substitution according to Conant-Stadler and Atchley protocols. Finally, using BUSTED and choosing AM494934 as foreground branch (Fig. 8a), the amino acid interval 205-224 underwent positive selection where constrained evidence ratio was higher than optimized null evidence ratio (Fig. 8b). While the interval of constrained evidence ratio ranged from 2.4769 (codon 220) to 18.3480 (codon 219), the interval of optimized null evidence of ratio ranged between -1.6619 (codon 220) and 4.1607 (codon 206). In spite of providing congruent results with REL indicating that part of the dimerization region (Fig. 9) in the CP gene of AM494934 accession was under positive selection, by applying a threshold, BUSTED revealed no statistically significant episodic diversifying selection in the wide-gene. Interestingly, BUSTED confirmed that amongst SW6 phylogroup components, two protruding branches of the tree had the highest expected substitution sites, those of AM494934, and AF332614 accessions (Fig. 8a) as shown earlier with network split tree (Fig. 3).

## 4 Discussion

In 2008, Fiore and colleagues characterized two Chilean peach isolates of PNRSV whose CP sequences were deposited in GenBank under the accession numbers EF565255, and EF565256. They noted that both of them were phylogenetically closely related to a previously described cherry isolate by Scott et al. [52] having the accession number AF013287. Fiore et al. [10] recognized that this group was outside of the classical phylogenetic scheme represented by PV96, PV32, and PE5 clusters; but at the same time, they stated that, based on the use of an alternative statistical approach, the four-cluster likelihood-mapping method, these accessions were considered as part of PV96 phylogroup. Gradually, the number of accessions increased including those Indian accessions described by Chandel et al. [53-55]. Based on a careful phylogenetic study of the coat protein-encoding gene of PNRSV, Boulila [15] stressed that a distinct phylogroup emerged due possibly to different evolutionary traits. Later on, Boulila et al. [16] suggested that members of this phylogroup should be assembled under the new appellation SW6. Moreover, Boulila [56] paid attention to the possibility that recombination may play a role in PNRSV evolution and his investigations demonstrated that the accession AF332614 (member of SW6 group) was a recombinant. Similarly, Boulila et al. [16] reported that a few Tunisian isolates underwent

recombination which classified them as outgroup. Here, the author sought for the occurrence of putative recombination events in the CP gene of different accessions belonging to SW6 cluster and emphasized their impact on selective pressure and evolutionary features of PNRSV. Recombination relevance and reliability necessitates an appropriate choice of different computer tools despite differences in their performance. In this study, six different programs (Dnasp, Bellerophon, SimPlot, SBP, GARD, and RDP4.70) were used and all were successful in detecting breakpoints in nine aligned sequences of members of SW6 phylogroup. Except the accessions AM408909, and AM408910, all remaining components of SW6 cluster were involved in the exchange of segments. Nowadays, the biggest challenge is inherent to globalization and the international trade in agricultural produce which broke down the traditional geographical barriers to the movement of viruses. Inadvertent introduction of planting material such as cuttings or seeds to a particular region or country may result in the dissemination of PNRSV. This may explain that the Chilean progenies i.e., EF565255, and EF565256 would have generated by the exchange of genetic material between American (AF013287), Indian (AM494934, AM920668, AM419814), and polish (AF332614) germplasms. RDP package and SimPlot methods indicated that AM494934 had two potential parents: AM920668, and EF565255. According to RDP4.70, the recombining region in the CP of AM494934 accession was located in the C-terminal zone of CP whose part of it is required for dimer formation [57]. Recombination gave rise to a new segment having the sequence: VNSRPPAGRPYGRAPNALDA (position 205-224) (Fig. 9). Consequently, part of dimerization region (DR) of PNRSV CP (VNSRPPAGRPY) (position 205-215) underwent recombination. More importantly, it would be interesting to search for its impact on CP-CP interaction. Figure 9 highlights that CP dimerization region (DR) was less conserved than RNA-binding domain (RBD). The PNRSV CP has the capacity to bind to the 3'UTR of its RNA3 and 4 [58, 59] through an RNA-binding domain (RBD) rich in R residues located at the N-terminal region between amino acid residues 25 to 50 of the protein [59, 60]. RBD is necessary for different viral processes, e.g. genome activation, RNA replication and translation [61]. Phylogenetic trees are an essential tool to identify and characterize recombination events in viruses. Evolutionary history is classically represented by a strictly bifurcating tree, which means that once two branches are formed, they subsequently do not interact with each other. When conflicting signals such as horizontal gene transfer or recombination occur, reticulations among phylogenetic tree branches can result. Using NeighborNet algorithm implemented in SplitsTree4 software [21], networked relationships among analyzed accessions sequences, were noted. These incompatibilities represented by boxes meant that putative recombination events occurred which, by the way, were evidenced by the different recombination detection methods summarized above. Recombination may therefore explain that members of SW6 evolved differently and constituted a clearly distinct phylogroup (Fig. 10). Moreover, all components of the three remaining phylogroups, i.e., PV32, PV96, and PE5 were tested in order to detect possible recombination. None of them has proven recombinant making SW6 as the unique recombinant group. Furthermore, the recombination rate can be influenced by selection pressure exerted on the size and stability of the genome. In fact, recombination tends to purge the genome of accumulated deleterious mutations as well as to create a mosaic structure by dispersing beneficial combinations of mutations. In agreement with neutrality tests and SNAP algorithm output, most of HyPhy models used in this study pointed out that SW6 phylogroup members possessed CP genes which underwent predominantly purifying selection, i.e., SLAC, FUBAR, FEL, IFEL, and PARRIS. In contrast, REL, MEME and BUSTED models showed that adaptive selection occurred in the C-terminal tail of the CP (including part of dimerization region) (Fig. 9). The genesis of AM494934 accession segment was probably a result of recombination as revealed by RDP4.70 algorithm.

## References

1. E.P. Rybicki, "Bromoviridae. In the sixth report of the International Committee on Taxonomy of Viruses,". Eds, F.A. Murphy, C.M. Fauquet, D.H.L. Bishop, S.A. Ghabrial, A.W. Jarvis, G.P. Martelli, M.A. Mayo, and M.D. Summers (eds.) , Springer-Verlag, Vienna, Austria, pp. 405-457, 1995.
2. A.M.Q. Adams M.J. King, E.B. Carstens, E.J. Lefkowitz, "Classification and nomenclature of viruses. In: Virus taxonomy,". Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, 2011.

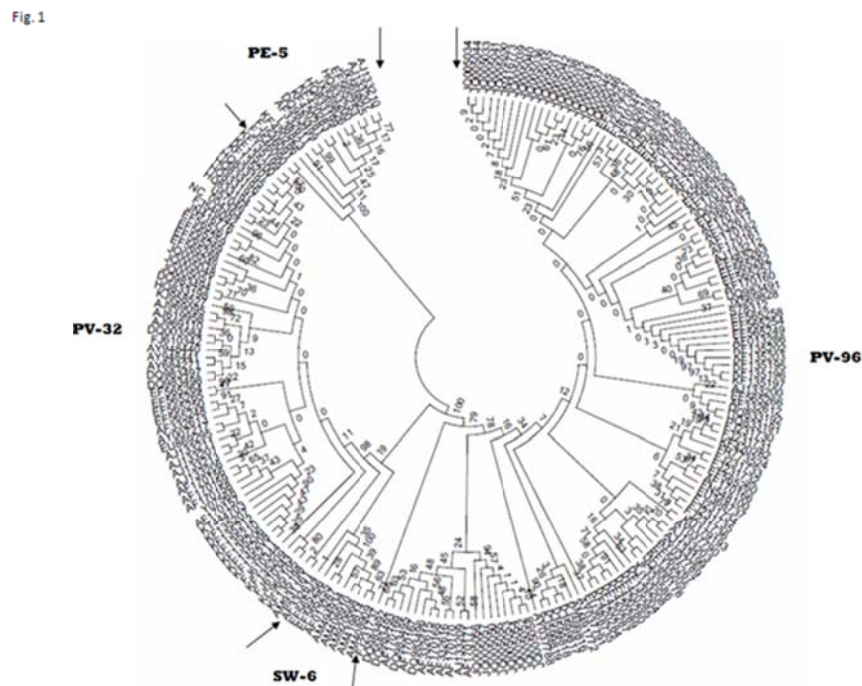
3. L. Van Vloten-Doting, R.I.B. Francki, R.W. Fulton, J.M. Kaper, L.C. Lane, "Tricornaviridae- a proposed family of plant viruses with tripartite single-stranded RNA genomes,". *Intervirology*, vol. 15, pp. 198-203, 1981.
4. E.J. Bachman, S.W. Scott, G.E. Xin, V.B. Vance, "The complete nucleotide sequence of Prune dwarf virus RNA3: implication for coat protein activation of genome replication in ilarviruses,". *Virology*, vol. 201, pp. 127-131, 1994.
5. D. Guo, E. Maiss, G. Adam, R. Casper, "Prunus necrotic ringspot ilarvirus: Nucleotide sequence of RNA 3 and the relationship to other ilarviruses based on coat protein comparison,". *J. Gen. Virol.*, vol. 76, pp. 1073-1079, 1995.
6. R.W. Hammond, J.M. Crosslin, "The complete nucleotide sequence of RNA 3 of a peach isolate of Prunus necrotic ringspot virus,". *Virology*, vol. 208, pp. 349-353, 1995.
7. S.W. Scott, X. Ge, "The complete nucleotide sequence of RNA 3 of Citrus leaf rugose and Citrus variegation ilarviruses,". *J. Gen. Virol.*, vol. 76, pp. 957-963, 1995.
8. D. Vaskova, K. Petrzik, R. Karesova, "Variability and molecular typing of the woody-tree infecting Prunus necrotic ringspot ilarvirus,". *Arch. Virol.*, vol. 145, pp. 699-709, 2000.
9. F. M. Codoñer, J. M. Cuevas, J. A. Sanchez-Navarro, V. Pallas, S. F. Elena, "Molecular evolution of the plant virus family Bromoviridae based on RNA3-encoded proteins". *J. Mol. Evol.*, vol. 61, pp. 697-705, 2005.
10. N. Fiore, T.V.M. Fajardo, S. Prodan, M.C. Herranz, F. Aparicio, J. Montealegre, S.F. Elena, V. Pallas, J. Sanchez-Navarro, "Genetic diversity of the movement and coat protein genes of South American isolates of Prunus necrotic ringspot virus,". *Arch. Virol.*, vol. 153, no. 5, pp. 909-919, 2008.
11. J.E. Oliver, J. Freer., R.L. Andersen, K.D. Cox, T.L. Robinson, M. Fuchs, "Genetic diversity of Prunus necrotic ringspot virus isolates within a cherry orchard in New York,". *Plant Dis.*, vol. 93, pp. 599-606, 2009.
12. F. Aparicio, A. Myrta, B. Di Terlizzi, V. Pallas, "Molecular variability among isolates of Prunus necrotic ringspot virus from different Prunus spp. *Phytopathology*, vol. 89, pp. 991-999, 1999.
13. F. Aparicio, V. Pallás "Molecular variability analysis of the RNA 3 of fifteen isolates of Prunus necrotic ringspot virus sheds light on the minimal requirements for the synthesis of the subgenomic RNA,". *Virus Genes*, vol. 25, pp. 75-84, 2002.
14. M. Glasa, E. Betinová, O. Kúdela, Z. Subr, "Biological and molecular characterization of Prunus necrotic ringspot virus isolates and possible approaches to their phylogenetic typing,". *Ann. App. Biol.*, vol. 140, pp. 279-283, 2002.
15. M. Boulila, "Putative recombination events and evolutionary history of five economically important viruses of fruit trees based on the coat protein-encoding gene sequence analysis,". *Biochem. Genet.* vol. 48, pp. 357-375, 2010.
16. M. Boulila, S. Ben Tiba, S. Jilani, "Molecular adaptation within the coat-protein-encoding gene of Tunisian almond isolates of Prunus necrotic ringspot virus,". *J. Gen.*, vol. no. 921, pp. 11-24, 2013.
17. M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, "Clustal W and Clustal X version 2.0,". *Bioinformatics*, vol. 23, pp. 2947-2948, 2007.
18. F. Corpet, "Multiple sequence alignment with hierarchical clustering,". *Nucl. Acids Res.*, vol. 16, pp. 10881-10890, 1988.
19. S. Kumar, G. Stecher, K. Tamura, "MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets,". *Mol. Biol. Evol.*, vol. 33, no. 7, pp. 1870-1874, 2016.
20. M. Kimura, "A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences,". *J. Mol. Evol.* 16, pp. 111-120, 1980.
21. D.H. Huson, D. Brayant, "Application of phylogenetic networks in evolutionary studies,". *Mol. Biol. Evol.*, vol. 23, no. 2, pp. 254-267, 2006.
22. K.S. Lole, R.C. Bollinger, R.S. Paranjape, D. Gadkari, S.S. Kulkarni, N.G. Novak, R. Ingersoll, H.W. Sheppard, S.C. Ray, "Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination,". *J. Virol.*, vol. 73, no. 1, pp. 152-160.
23. D.P. Martin, B. Murrell, M. Golden, A. Khoosal, B. Muhire, "RDP4: Detection and analysis of recombination patterns in virus genomes". *Virus Evol.* vol 1, no. 1, pp. 1-5, 2015.
24. D. Martin, E. Rybicki, RDP: "detection of recombination amongst aligned sequences,". *Bioinformatics*, vol. 16, pp. 562-563, 2000



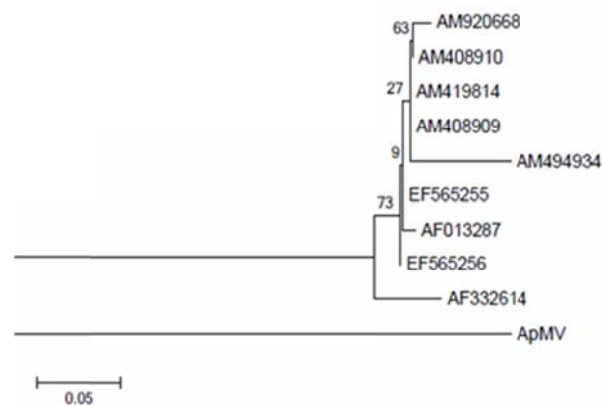
- 25.M. Padidam, S. Sawyer, C.M. Fauquet, "Possible emergence of new geminiviruses by frequent recombination,". *Virology*, vol. 265, pp. 218-225. 1999.
- 26.D.P. Martin, D. Posada, K.A. Crandall, C. Williamson, "A modified bootscan algorithm for automated identification of recombination sequences and recombination breakpoints,". *AIDS Res. Hum Retrov.*, vol. 21, pp. 98-102. 2005.
- 27.J.M. Smith, "Analyzing the mosaic structure of genes,". *J. Mol. Evol.*, vol. 34, pp. 126-129, 1992.
- 28.D. Posada, K. "Crandall Evaluation of methods for detecting recombination from DNA sequences: Computer simulation". *Proc. Nat. Acad. Sci.* vol. 98, pp. 13757-13762, 2001.
- 29.M.J. Gibbs, J.S. Armstrong, A.J. Gibbs, "Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences,". *Bioinformatics*, vol. 16, pp. 573-582, 2000.
- 30.M.F. Boni, D. Posada, M.W. Feldman, "An exact nonparametric method for inferring mosaic structure in sequence triplets,". *Genetics*, vol. 176, pp. 1035-1047, 2007.
- 31.T. Huber, G. Faulkner, P. Hugenholtz, "Bellerophon: A program to detect chimeric sequences in multiple sequence alignments,". *Bioinformatics*, vol. 20, 2317-2319, 2004.
- 32.P. Librado, J. Rozas, DnaSP v5: "A software for comprehensive analysis of DNA polymorphism data,"". *Bioinformatics*, vol. 25, pp. 1451-1452, 2009.
- 33.S.L. Kosakovsky Pond, D. Posada, M.B. Gravenor, C.H. Woelk, S.D. Frost, "GARD: a genetic algorithm for recombination detection,". *Bioinformatics*, vol. 22, no. 24, pp. 3096-3098, 2006a.
- 34.S.L. Kosakovsky Pond, D. Posada, M.B. Gravenor, C.H. Woelk, S.D. Frost, "Automated phylogenetic detection of recombination using a genetic algorithm,". *Mol. Biol. Evol.* vol. 23, pp. 1891-1901, 2006b.
- 35.H. Akaike, "A new look at the statistical model identification". *IEEE Trans. Autom. AC*, vol. 19, pp. 716-723, 1974.
- 36.H. Kishino, M. Hasegawa "Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea,". *J. Mol. Evol.* vol. 29, pp. 170-179, 1989.
- 37.S.L. Kosakovsky Pond, S.D.W. Frost, S.V. Muse, "HyPhy: hypothesis testing using phylogenies,". *Bioinformatics*, vol. 21, pp. 676-679, 2005.
- 38.F. Tajima "Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism,". *Genetics*, vol. 123, pp. 585-595, 1989.
- 39.Y. .X, Fu, W.H. Li, "Statistical tests of neutrality of mutations". *Genetics*, vol. 133, pp. 639-709, 1993
- 40.G.A. Watterson, "On the number of segregating sites in general models without recombination". *Theor. Popul. Biol.*, vol. 7, pp. 256-276, 1975
- 41.B. Korber, "HIV signatures and similarities. In: Computational and Evolutionary Analysis of HIV Molecular Sequences", A.G. Rodrigo and G.H. Learn Jr. Dordrecht, The Netherlands: Kluwer Academic Publishers Ed., pp. 55-72, 2000.
- 42.S.L. Kosakovsky Pond, S.D.W. Frost, "Datamonkey: rapid detection of selective pressure on individual sites of codon alignments, ". *Bioinformatics*, vol. 21, pp. 2531-2533. 2005a.
- 43.S.L. Kosakovsky Pond, S.D.W. Frost, "A genetic algorithm approach to detecting lineage-specific variation in selection pressure,". *Mol. Biol. Evol.*, vol. 22, pp. 478-485. 2005b.
- 44.S.L. Kosakovsky Pond, S.D. Frost, Z. Grossman, M.B. Gravenor, D.D. Richman, A.J. Brown, "Adaptation to different human populations by HIV-1 revealed by codon-based analyses,". *PLOS Comput. Biol.*, vol. 2, pp. 530-538, 2006c.
- 45.B. Murrell, J.O. Wertheim, S. Moola, T. Weighill, K. Scheffler, S.L. Kosakovsky Pond, Detecting individual sites subject to episodic diversifying selection. *PLOS Genetics* 8: e1002764. doi:10.1371/journal.pgen.1002764. 2012.
- 46.B. Murrell, S. Moola, A Mabona., T. Weighill, D. Sheward, S.L. Kosakovsky Pond, K. Scheffler, FUBAR: A Fast, Unconstrained Bayesian AppRoximation for inferring selection. *Mol. Biol. Evol.*, vol. 30, pp. 1196-1205, 2013.
- 47.K. Scheffler, D.P. Martin, C. Seoighe "Robust inference of positive selection from recombining coding sequences,". *Bioinformatics*, vol. 22, pp. 2493-2499, 2006.
- 48.W. Delpont, A.F.Y. Poon, S.D.W. Frost, S.L. Kosakovsky Pond, "Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology,". *Bioinformatics*, vol. 29, pp. 2455-2457, 2010.
- 49.G.C. Conant, G.P. Wagner, P.F. Stadler "Modeling amino acid substitution pattern in orthologous genes,". *Mol. Phyl. Evol.*, vol. 42, no. 2, pp. 298-307, 2007.

- 50.W.R. Atchley, J. Zhao, A.D. Fernandes, T. Druke "Solving the protein sequence metric problem,". PNAS, vol. 102, no. 18, pp. 6395-6400, 2005.
- 51.B. Murrell, S. Weaver, M. D. Smith, J. O. Wertheim, S. Murrell, A. Aylward, K. Eren, T. Pollner, D. P. Martin, D. M. Smith, K. Scheffler, S. L. Kosakovsky Pond, "Gene-Wide Identification of Episodic Selection". Mol. Biol. Evol., vol. 32, no. 5, pp. 1365–1371, 2015.
- 52.S.W. Scott, M.T. Zimmerman, Xin Ge, D.J. MacKenzie, "The coat protein and putative movement proteins of isolates of Prunus necrotic ringspot virus from different host species and geographic origins are extensively conserved,". Europ. J. Plant Pathol., vol. 104, pp. 155–161, 1998.
- 53.V. Chandel, T. Rana, V. Hallan, A.A. Zaidi, "Occurrence of Prunus necrotic ringspot virus on nectarine (*Prunus persica*) in India,". Europ. Plant Prot. Org. (EPPO), vol. 38, pp. 223-225, 2008a.
- 54.V. Chandel, T. Rana, A. Handa, P.D. Thakur, V. Hallan, A.A. Zaidi, "Incidence of Prunus necrotic ring spot virus on *Malus domestica* in India,". J. Phytopathol., vol. 156, pp. 382-384, 2008b.
- 55.V. Chandel, T. Rana, V. Hallan, A.A. Zaidi, "Detection of Prunus necrotic ringspot virus in plum, cherry and almond by serological and molecular techniques from India,". Arch. Pytopathol. Plant Prot., vol. 44, no. 18, pp. 1779–1784, 2011.
- 56.M. Boulila "Molecular evidence for recombination in Prunus necrotic ringspot virus,". Plant Mol. Biol. Rep., vol. 27, pp. 189-198, 2009.
- 57.F. Aparicio, J. A. Sánchez-Navarro, V. Pallás, "In vitro and in vivo mapping of the Prunus necrotic ringspot virus coat protein C-terminal dimerization domain by bimolecular fluorescence complementation". J. Gen. Virol., vol. 87, pp. 1745-1750, 2006.
- 58.V. Pallás, J. A. Sanchez-Navarro, J. Diez, "In vitro evidence for RNA binding properties of the coat protein of Prunus necrotic ringspot ilarvirus and their comparison to related and unrelated viruses,". Arch. Virol., vol. 144, pp. 797-803. 1999.
- 59.F. Aparicio, M. Vilar, E. Perez-Paya, V. Pallas,"The coat protein of Prunus necrotic ringspot virus specifically binds to and regulates the conformation of its genomic RNA,". Virology, vol. 313, pp. 213-223, 2003.
- 60.V. Pallas, F. Aparicio, M. C. Herranz, K. Amari, M. A. Sanchez-Pina, A. Myrta, J. A. Sanchez-Navarro Iilarviruses of Prunus spp., "A continued concern for fruit trees,". Phytopathology, vol. 102, pp. 1108-1120, 2012.
- 61.J. F. Bol, "Replication of alfamo- and ilarviruses: Role of the coat protein,". Annu. Rev. Phytopathol., vol. 43, pp. 39-62, 2005.
- 62.R.H. Alrefai, P.J. Shiel, L.L. Domier, C.J. D'Arcy, P.H. Berger, S.S. Korban, "The nucleotide sequence of Apple mosaic virus coat protein gene has no similarity with other Bromoviridae coat protein genes,". J. Gen. Virol., vol. 75, no. 10, pp. 2847-2850, 1994.
- 63.T. Malinowski B. Komorowska, "Metoda powielania genu białka płaszczka (CP) wirusaneotropycznej plamistości pierścieniowej drzew pestkowych (PNRSV) i jej zastosowanie. Polish". Conf. Fruit Plants Prot., Skierniewice, Poland, pp.203, 1998.

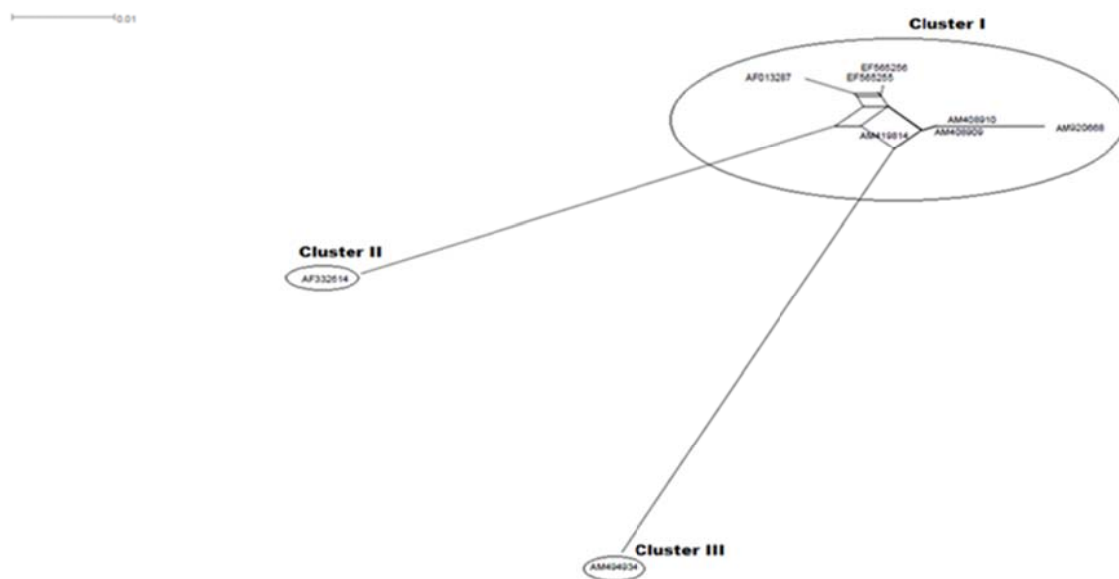
## Appendix A:



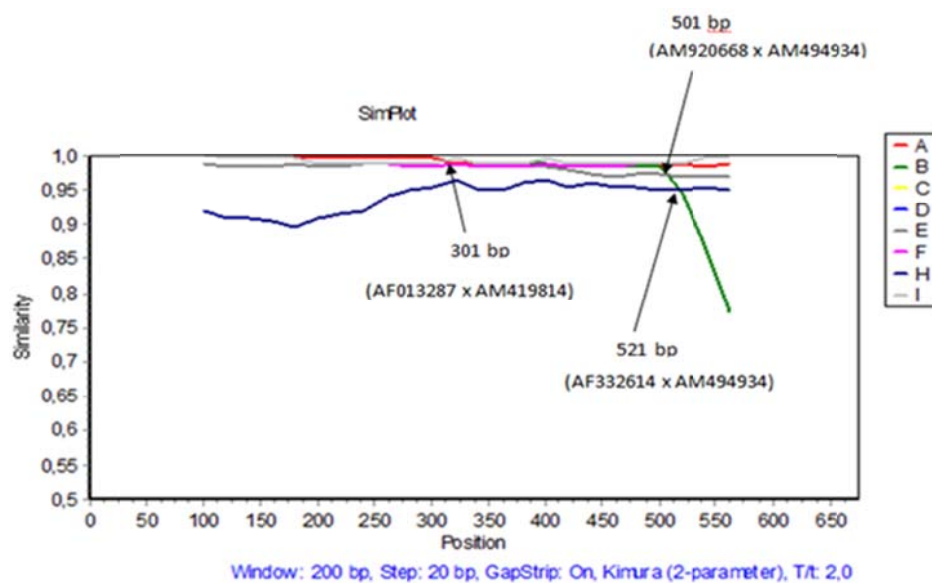
**Figure 1.** Based on CP analysis, the circle form of a phylogenetic tree involving 205 PNRSV accessions retrieved from GenBank and produced by the ML algorithm under assumption of the Model K2+G an option of MEGA7 software (Kumar *et al.* [19]). Four major groups were delineated including SW6. Bootstrap analysis was performed with 1,000 replicates. The numbers above the branches indicate the bootstrap confidence value.



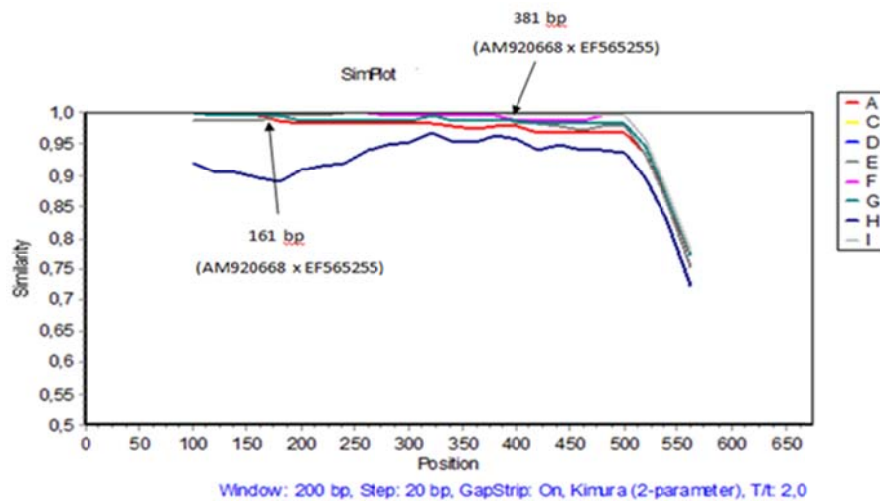
**Figure 2.** Dendrogram depicting the phylogenetic relationship among nine isolates of PNRSV forming SW6 phylogroup based on the nucleotide sequence of the CP domain. The tree was reconstructed using the ML algorithm under assumption of the Model K2 of the MEGA software. Bootstrap analysis was performed with 1,000 replicates. The numbers above the branches indicate the bootstrap confidence value. The scale bar shows the number of substitution per nucleotide. The CP sequence of an isolate of Apple mosaic virus (accession number: L03726) (Alrefai *et al.* [62]), was used as outgroup.



**Figure 3.** Neighbor Net for nine CP sequences of PNRSV isolates. The networked relationships are shown to indicate the presence of reticulate events. Boxes imply the likelihood of recombination. Three major clusters were delineated. The phylogenetic network was constructed with Splits Tree4 program (Huson and Bravant[21]). The scale bar shows the number of substitution per nucleotide.



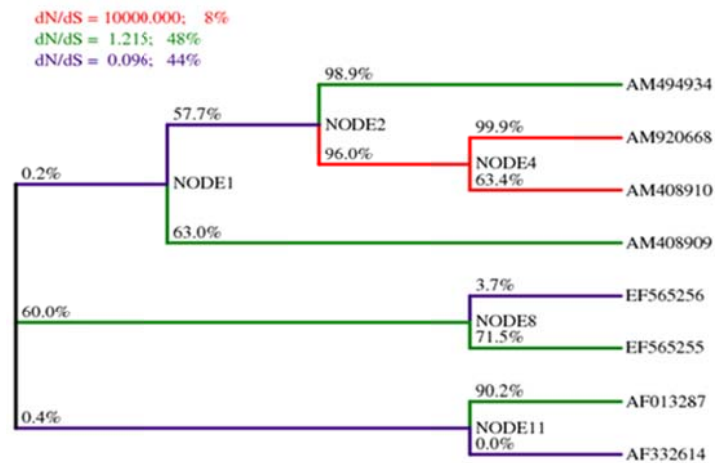
**Figure 4.** Similarity plot generated in SimPlot using a query sequence (applied to both EF565255 (G) and EF565256 (C) accessions) and all remaining sequences. Possible recombination sites were identified where sequence crossover occurred as indicated by arrows. The putative parentals are mentioned in brackets. Uppercase letters represent: A: AF013287, B: AM494934, C: EF565256, D: AM920668, F: AM408910, G: EF565255, H: AF332614, I: AM419814.



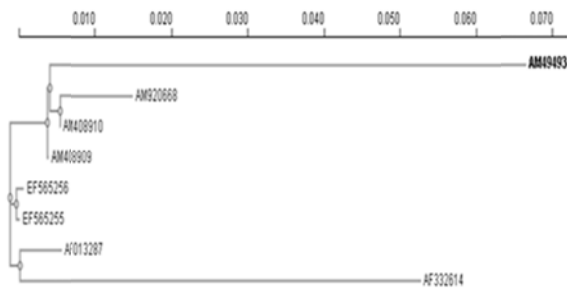
**Figure 5.** Similarity plot generated in SimPlot using a sequence (applied to AM494934 (B) accession) and all remaining sequences. Possible recombination sites were identified where sequence crossover occurred as indicated by arrows. The putative parents are mentioned in brackets. Uppercase letters represents: A: AF013287, B: AM494934, C: EF565256, D:AM408909, E:AM920668, F:AM408910, G: EF565255, H: AF332614, I: AM419814.



**Figure 6.** Schematic CP sequence display of nine accessions of PNRSV forming SW6 phylogroup and highlighting that AM494934 is a recombinant according to RDP4.70 software.



**Figure 7.** Lineage specific analysis of selective pressure in CP gene of PNRSV. A cladogram is shown with Maximum-Likelihood Estimates of lineage-specific  $dN/dS$  during CP gene evolution. Percentages for branch classes in the legend reflect the proportion of total tree length (measured in expected substitutions per site per unit time) evolving under the corresponding value of  $dN/dS$ . The branch classes are indicated by three different colors in inferred cladogram.



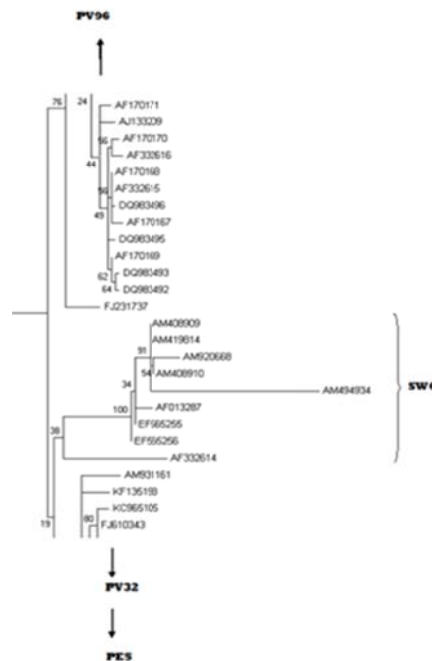
**Figure 8a.** BUSTED output showing possible foreground branches. Besides, AM494934 and AF332614 accessions had highest expected substitutions per site. The scale represents the expected substitutions per site.



**Figure 8b.** Taken as a foreground branch (while all remaining were background branches), AM494934 capsid protein was under positive selection in its C-terminal region where the constrained ratio (in green) was higher than optimized null evidence ratio (in red).



**Figure 9.** Amino acid sequence alignment of coat protein of nine accessions composing SW6 phylogroup. The basic region (RBD) carrying critical residues (in red) for the RNA binding of the CP is underlined and shaded in grey. The C-terminal region required for dimer formation (DR) is underlined too (not shaded). In addition, the sequence of the exchanged fragment by recombination according to RDP4.70 is written in red. It is worth noting that part of the recombined segment belongs to the DR region. Numbers at the top represent the corresponding residues in the full-length protein. Dots indicate identical residues.



**Figure 10.** Dendrogram pointing out SW6 is a distinct phylogroup offset from the classical scheme PV96-PV32-PE5. The tree was using the ML algorithm under assumption of the Model K2 of the MEGA7 software. Bootstrap analysis was performed with 1,000 replicates. The numbers above the branches indicate the bootstrap confidence value.

## Appendix B:

**Table 1.** Characteristics of *prunus necrotic ringspot virus* (PNRSV) accessions used in this study.

Accessions	Isolate	Host	Country	Authors
AF013287	SW6	Sweet cherry	USA	Scott <i>et al.</i> [51]
AM494934	/	Plum	India H.P.	Chandel <i>et al.</i> [54]
AM920668	/	Cherry	India: Palampur	Chandel <i>et al.</i> [54]
AM408909	/	Nectarine	India : Kullu	Chandel <i>et al.</i> [52]
AM408910	/	<i>Prunus persica</i> var. nucipersica	India: Kullu	Chandel <i>et al.</i> [52]
AM419814	<i>Golden delicious</i>	Malus x domestica	India: Kullu	Chandel <i>et al.</i> [53]
EF565255	PchCl.loa1	Peach cv. Loadell	Chile	Fiore <i>et al.</i> [10]
EF565256	PchCl.loa2	Peach cv. Loadell	Chile	Fiore <i>et al.</i> [10]
AF332614	NRSiz5/ WWL 1/7	Plum	Poland	Malinowski and Komorowska [63]

**Table 2.** Percentage nucleotide identity among the PNRSV isolates examined. The highest and lowest rates are indicated in bold.

AM920668	100%								
AM494934	92.89%	100%							
AM408910	98.96%	93.93%	100%						
EF565256	98.07%	93.49%	99.11%	100%					
AM419814	98.81%	94.08%	99.85%	99.26%	100%				
EF565255	98.22%	93.49%	99.26%	99.85%	99.4%	100%			
AF013287	97.48%	92.75%	98.52%	99.11%	98.66%	99.26%	100%		
AF332614	92.89%	88.9%	93.93%	94.37%	94.08%	94.23%	94.08%	100%	
AM408909	98.81%	94.08%	99.85%	99.26%	100%	99.4%	98.66%	94.08%	100%
	AM920668	AM494934	AM408910	EF565256	AM419814	EF565255	AF013287	AF332614	AM408909



**Table 3.** Percentage amino acid identity (below the diagonal) and the similarity (above the diagonal) among the PNRSV isolates examined. The highest and lowest rates are indicated in bold.

<b>AF013287</b>	<b>100%</b>	91.55%	99.11%	99.11%	96.00%	98.66%	99.11%	96.44%	99.11%
<b>AM494934</b>	90.22%	<b>100%</b>	92.44%	92.44%	90.22%	92.00%	92.44%	88.88%	92.44%
<b>EF565256</b>	99.11%	99.11%	<b>100%</b>	100%	96.88%	99.55%	100%	96.44%	100%
<b>AM408909</b>	98.66%	91.55%	99.55%	<b>100%</b>	96.88%	99.55%	100%	96.44%	100%
<b>AM920668</b>	95.11%	88.88%	96.00%	96.44%	<b>100%</b>	97.33%	96.88%	97.33%	96.88%
<b>AM408910</b>	98.22%	91.11%	99.11%	99.55%	96.88%	<b>100%</b>	99.55%	96.00%	99.55%
<b>EF565255</b>	99.11%	91.11%	100%	99.55%	96.00%	99.11%	<b>100%</b>	96.44%	100%
<b>AF332614</b>	94.66%	87.11%	95.55%	95.55%	92.44%	95.11%	98.55%	<b>100%</b>	96.44%
<b>AM419814</b>	98.66%	91.55%	99.55%	100%	96.44%	99.55%	99.55%	95.55%	<b>100%</b>
	<b>AF013287</b>	<b>AM494934</b>	<b>EF565256</b>	<b>AM408909</b>	<b>AM920668</b>	<b>AM408910</b>	<b>EF565255</b>	<b>AF332614</b>	<b>AM419814</b>

**Table 4.** Population genetic parameters and neutrality tests calculated for the CP gene.  $M$  = number of sequences,  $S$  = number of segregating sites,  $\theta$  = The statistic  $\theta$  from the number of segregation sites ( $S$ ) (Watterson  $\theta$  estimator, [40]) and the average of synonymous and non-synonymous substitutions,  $\pi$  = nucleotide diversity (estimated by the average number of nucleotide differences between two random sequences in a population),  $Hd$  = haplotype diversity,  $\Delta Hd$  = the variance of haplotype diversity,  $K$  = Average of number of pairwise nucleotide differences, +Tajima's  $D$  and Fu and Li's  $D$  and  $F$  tests measure the departure from neutrality for all mutation in CP cistron. \*Average of all pairwise comparisons (<http://www.hiv.lanl.gov/cgi-bin/SNAP/WEBSNAP/SNAP.cgi>).

$M$	Population statistics						Test of neutrality +			Synonymous and non-synonymous statistics after SNAP algorithm *		
	$S$	$\theta$	$\pi$	$Hd$	$\Delta Hd$	$K$	Tajima's statistics	Fu and Li's $F$ statistics	Fu and Li's $D$ statistics	$dN$	$dS$	$dS/dN$
9	86	0.04913	0.03243	0.972	0.00409	21.861	-1.75204	-2.12993	-1.94133	0.0282	0.0810	2.8723

**Table 5.** Negatively selected sites in CP gene estimated by (i) SLAC (a), IFEL (b), and FEL (d) Models where codon position, normalized  $dN-dS$  [ $(dN-dS)/(\text{codon tree length})$ ], and  $P$ -value were estimated, (ii) FUBAR Model (c) where the means of posterior distribution of synonymous ( $\alpha$ ) and non-synonymous ( $\beta$ ) substitution rates over sites as well as the mean posterior probability for  $\omega=\beta/\alpha < 1$  at a site (pervasive purifying selection), and  $\omega=\beta/\alpha > 1$  (pervasive diversifying selection), were estimated, (iii) PARRIS Model (e) where inferred rate distribution (synonymous rate and  $\omega$  ratio) for the null (M1) and the alternative (M2) models, were determined. Positively selected sites were estimated by (i) REL Model (f) where codon position, Elevated  $dN$  and  $dS$ , normalized  $dN-dS$  [ $(dN-dS)/(\text{codon tree length})$ ], posterior probability and Bayes factors were calculated, (ii) MEME Model (g) where the distribution of synonymous ( $\alpha$ ) and non-synonymous ( $\beta$ ) substitution rates over sites inferred by the model where the proportion of branches with  $\beta > \alpha$  is significantly greater than 0, were determined. The  $P$ -value is derived using a mixture of  $\chi^2$  distribution and  $q$ -values using simes's procedure which controls the false discovery rate under the strict neutral null (likely to be conservative).

**a**

SLAC Model	Codon	$dN-dS$	Normalized $dN-dS$	$p$ -value
Negatively selected site	98	-0.977961	-6.38739	0.0592723

**b**

IFEL Model	Codon	$dN$	$dS$	$dS/dN$	Normalized $dN-dS$	$p$ -value
Negatively selected site	219	0	10000	0.000	-65456.8	0.0137778

**c**

FUBAR Model	Codon	$\alpha$	$\beta$	$\beta-\alpha$	Posterior probability $\beta < \alpha$
Negatively selected site	98	6.07395	0.133044	-5.9409	0.986734
	167	4.43708	0.127438	-4.30964	0.923782
	219	8.9256	1.56947	-7.35612	0.921252

**d**

FEL Model	Codon	$dS$	$dN$	$dN/dS$	Normalized $dN-dS$	$p$ -value
Negatively Selected sites	21	12.4842	0	0.000	-81.7174	0.036957
	30	7.2056	0	0.000	-47.1655	0.0675461
	33	10.7667	0	0.000	-70.4751	0.044143
	70	12.4842	0	0.000	-81.7173	0.0383188
	98	45.9032	0	0.000	-300.467	0.00367723
	111	12.4842	0	0.000	-81.7173	0.0383188
	145	12.4842	0	0.000	-81.7173	0.0383188
	151	6.70809	0	0.000	-43.909	0.0934671
	153	6.25732	0	0.000	-40.9584	0.0907652
	154	10.7667	0	0.000	-70.4751	0.0458604
	167	64.7523	0	0.000	-423.848	0.0102524
	170	83.6952	0	0.000	-547.842	0.0200195
	183	18.2176	0	0.000	-119.247	0.0267242
	219	10000	4.02261	0.000	-65430.4	0.00627941
	222	13.0772	0	0.000	-85.599	0.0502493

**e**

PARRIS	Coding region	Inferred Rate distributions	Synonymous rate			$\omega$ ( $dN/dS$ ) ratio				
			Rate Class	1	2	3	Rate class	1	2	3
CP gene	Null Model (M1)	$dS$		0.26	1.19	2.55	$\omega$	0.00	1.00	/
		Probability		0.676	0.000	0.324	Probability	0.553	0.447	/
	Alternative Model (M2)	$dS$		0.26	1.19	2.55	$\omega$	0.00	1.00	2.90
		Probability		0.676	0.000	0.324	Probability	0.553	0.447	0.000

f

REL Model	Codon	E( <i>dS</i> )	E ( <i>dN</i> )	Normalized E( <i>dN-dS</i> )	Posterior Probability	Bayes Factor
Positively	47	0.605251	1.37385	0.768594	0.985182	244.808
Selected	206	0.635463	1.37582	0.740359	0.976035	149.969
site	207	0.646258	1.37436	0.728106	0.961149	91.0961
	208	0.602817	1.37396	0.771147	0.985477	249.863
	210	0.61788	1.37338	0.755502	0.982833	210.816
	212	0.600499	1.37378	0.773285	0.985496	250.187
	213	1.09023	1.37563	0.285404	0.956183	80.3543
	214	0.699109	1.37405	0.674939	0.956262	80.5056
	215	0.625616	1.37314	0.747527	0.979679	177.523
	216	0.728853	1.37403	0.64518	0.954607	77.4365
	217	0.662233	1.37545	0.713215	0.970837	122.58
	221	0.726155	1.37448	0.648324	0.959585	87.4268
	224	0.873918	1.37362	0.499704	0.945871	64.3444

g

MEME	Codon	$\alpha$	$\beta^-$	Pr[ $\beta=\beta^-$ ]	$\beta^+$	Pr[ $\beta=\beta^+$ ]	<i>p</i> -value	<i>q</i> -value
	124	0.00040016	6.40256.10 <sup>-5</sup>	0.86477	226.272	0.13523	0.0556043	1
	206	0	0	0.753908	40.9676	0.246092	0.0630978	1