Sequence diversity and potential recombination events in the coat protein
gene of Apple stem pitting virus
Beata Komorowska¹, Paweł Siedlecki², Szymon Kaczanowski³, Beata Hasiów-Jaroszewska⁴ and Tadeusz Malinowski¹

¹Department of Plant Protection, Research Institute of Horticulture, Konstytucji 3
Maja 1/3, 96-100 Skierniewice, Poland
² Department of Bioinformatics, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawińskiego 5a, 02-106 Warsaw, Poland
³Department of Plant Molecular Biology, Institute of Experimental Plant Biology,
University of Warsaw, Miecznikowa 1, 02-096 Warsaw, Poland
⁴Department of Virology and Bacteriology, Institute of Plant Protection-National
Research Institute, Wędorka 20, 60-318 Poznań, Poland

SUMMARY

The variability of the *Apple stem pitting virus* (ASPV) coat protein (CP) gene was investigated. The CP gene of ten virus isolates from apple and pear trees was sequenced. Comparison of all sequenced virus isolates revealed high diversity of the CP gene (70.7-93.5 % at the nucleotide level and 77.8-98.7 % at the amino acid level). Additionally, one or two deletions in the N-terminal part of the coat protein gene of the studied virus isolates were identified. The ratio of nonsynonymous to synonymous polymorphic sites indicated that purifying selection has acted to eliminate deleterious mutations in coding sites. Moreover, the evidences for recombination in analyzed sequences were provided. It is likely that recombination, along with selection, enhances the speed of elimination of deleterious mutations in ASPV, following the mutational deterministic hypothesis of Kondrashov.

Key words: ASPV, sequence diversity, selection, recombination

*Apple stem pitting virus* (ASPV), a member of *Foveavirus* genus in the family *Betaflexiviridae* (Martelli et al., 2007), is one of the most common viruses of apple and pear trees worldwide. ASPV causes xylem pits in the stem of *Malus pumila* Virginia Crab as well as epinasty and decline of *Malus domestica* Spy 227, but remains symptomless in apple cultivars (Nemeth, 1986). This virus is also identified as a causal agent of pear vein...
yellows disease (Schwarz and Jelkmann, 1998). There is no known insect vector for this
virus (Martelli and Jelkmann, 1998). The ASPV is a filamentous virus with the positive-
sense, single-stranded RNA forming 15 nm wide and 800 nm long particles (Koganezawa
and Yanase, 1990). The previous results of the genetic diversity of ASPV isolates showed
that nucleic acid identity among coat protein (CP) gene sequences was found to be
between 81.7 %-99.3 % (Schwarz and Jelkmann, 1998) and 64.8 %-89.6 % (Yoshikawa et
al., 2000). Additionally, one or two deletions in the N-terminal part of the viral CP gene
were reported (Schwarz and Jelkmann, 1998; Yoshikawa et al., 2000). Although the
 genetic diversity of ASPV isolates has been studied, the knowledge of the evolutionary
history is still lacking. Viral coat proteins are presumed to have evolved more rapidly than
proteins involved in replication and expression of virus genomes (Callaway et al., 2001).
For this reason, the coat protein genes have been used in the phylogenetic comparisons of
different viruses (Koonin and Gorbalenya, 1989; Zimmern, 1988) as well as in this work.
The recombination events could play an important role in deleterious mutation
clearance. Such a function of recombination was firstly suggested by Muller (1964) who
described a phenomenon termed Muller’s ratchet. Muller proposed that in a finite
population the best haplotypes might be lost by chance of through recurrent deleterious
mutations and without recombination they could not be recreated (Iles et al., 2003). Also
according to the modern mutational deterministic hypothesis of Kondrashov (Kondrashov,
1988), recombination is used for deleterious mutation clearance. It should be noted here
that some evolutionary biologists use the term “sex” in case of viruses, which is
synonymous to genetic exchange. See for example work done by Turner group [Orig Life
Evol Biosph. 2003 33:95-108 Turner PE] which studied experimental evolution of sex
using viruses as a model. Recombination in natural populations has been reported for
several plant viruses (Glasa et al., 2004; Ohshima et al., 2002; Tomitaka and Ohshima,
2006). To our knowledge there is no data on potential recombination between ASPV
isolates.
Sequence analysis of the ASPV CP gene has been conducted and the ratio of
synonymous and nonsynonymous polymorphisms was calculated. This work also
addresses the question what is the pattern of selection shaping the observed diversity of
ASPV with a coat protein gene as a model. Additionally, a computational analysis of
sequences to investigate the extent of recombination events as well as to
reconstruct the most likely evolutionary history of the recombinants in the population of
ASPV genomes has been described.
The virus isolates originated from apple and pear cultivar collection maintained at the Research Institute of Pomology and Floriculture, and from commercial orchards in Poland (Table 1). Isolate N1 was found in the indicator plant, but the original source of the virus was uncertain. Total nucleic acids were isolated from the leaves using the silica capture (SC) method described originally by Boom et al. (1990) and adapted to the diagnosis of plant viruses by Malinowski (1997). Amplification was performed using SuperScript One-Step RT-PCR kit (Invitrogen, Carlsbad CA, USA and the primers ASPB2-ASPR3CP (Komorowska et al., 1999). RT-PCR with primer pair ASPB2-77ASPB3 yielded the cDNA fragment encompassing viral CP gene of all studied virus isolates. Besides amplification of a 1443 bp product as calculated from the previously sequenced PA66 and PSA-H isolates (Jelkmann, 1994), bands of lower molecular weight were observed with other virus sources. The amplified RT-PCR products were ligated into the pCR 2.1-TOPO vector and used for transformation of TOP10 One-Shot E. coli competent cells according to the manufacturer’s instructions (Invitrogen, Carlsbad CA, USA). The cloned PCR fragments were sequenced either with universal primers M13 forward and M13 reverse or virus specific primers C8849 (MacKenzie et al., 1997) and 85ASPV9019 (Schwarz and Jelkmann, 1998) targeting the regions inside the CP gene. Sequencing was performed using an automated sequencer ABI PRISM 310 in the Institute of Biochemistry and Biophysics in Warsaw, Poland. All the sequences were read at least twice. Sequence data were analyzed using Lasergene v. 7.1 software package (DNASTAR, USA). Four samples from apple (MT32, MT24, J335, ST181) and one from an unknown origin (N1) showed an 1389 bp product. Two samples from pear (ST54, ST113, ST132) were 1323 bp long. PCR product of 1320 bp was amplified from pear GNKVII/34. The multiple alignments, including the previously reported sequences of isolates PA66 and PSA-H (Jelkmann, 1994), demonstrated that the observed differences in size of the RT-PCR products were due to deletions in the N-terminal part of the coat protein gene (Table 1). These results are in accordance with the outcomes obtained for ASPV isolates in other centers (Schwarz and Jelkmann, 1998; Yoshikawa et al., 2000). Moreover, the deletion starting from nt 8233(99)Table 1) was found only in the CP gene of virus pear isolates (ST113, ST132, GNKVII/34). Similar results have been shown previously for isolate I7 described by Schwarz and Jelkmann, 1998. There is also another ASPV pear isolate in GenBank (accession no. AY572458) with two deletions in CP gene. Thus, it seems that the second
deletion might be unique for pear isolates. On the other hand the number of the characterized ASPV isolates (25) is too small to formulate the general rule. Although plant viruses with RNA genomes have a high potential for genetic variations (Garcia-Arenal et al., 2003) there are only a few examples of large natural deletions in the CP gene (Seifers et al., 2005; Szathmary et al., 2009).

Following the reading of CP gene sequences for ten ASPV isolates, the comparison among them and with other virus sequences from GenBank (accession nos. D21828, D21829, AB045371, EU708018, EU314950, DQ003336, AY572458) was carried out. Multiple sequence alignments were performed using the online service CLUSTALW (Thompson et al., 1994). The multiple alignments revealed low identity between ten studied virus sequences ranging from 70.7 % to 93.5 % at nucleotide level and from 77.8 % to 98.7 % at amino acid level. Similar ranges of values were obtained when comparisons were made with corresponding sequences available in GenBank. The percentage of nucleic acid identities among the sequences of the CP gene of all ASPV isolates was found to be between 71 % and 99.8 %. Comparison of amino acid sequences showed higher identities 78.2-99.5 %. Similar high levels of CP gene diversity are described for other members of the family Betaflexiviridae (Alabi et al., 2010; Teycheney et al., 2005). The lack of a biological vector for transmission of ASPV probably makes the CP more tolerant to sequence variability than the capsid proteins of vector borne plant viruses subjected to greater purifying selection on amino acid changes (Garcia-Arenal et al., 2001; Rubio et al., 2001; Vives et al., 2002).

All phylogenetic analyses were conducted in Datamonkey (Kosakovsky Pond and Frost, 2005a) which is a front-end to HyPhy (Kosakovsky Pond et al., 2005). HKY85 was chosen as the nucleotide substitution bias model, as our data set was rather small. This model was used in all phylogenetic studies. The recognized deletions were discarded from the CP sequences in the following analyses.

It is well known that recombination can affect obtained values for omega and phylogenetic reconstruction (Silva and Edwards, 2009). Therefore, the ASPV CP genes were initially analyzed for evidence of recombination events. The GARD (genetic algorithm recombination detection) and the SBP (single breakpoint recombination) methodologies were used to determine if recombination events are present in the studied sequences (Kosakovsky Pond and Frost 2005b; Kosakovsky Pond et al., 2006). Both methods differently identify non-recombinant fragments in the alignment and reconstruct a phylogenetic tree for each of them. Each of the subsequent experiments was repeated...
both with GARD and SBP tree sets. The quality of evolutionary trees was measured using minimum AIC values. If AIC for non-recombinant tree is smaller than AIC for a tree with recombination event the delta AIC should be high. The SBP algorithm found evidence of recombination at position 604 (with a high delta AIC of 854). The GARD algorithm was able to find two possible recombination sites within ASPV capsid sequences in positions 618 (p=0.0008; high confidence, small delta AIC) and 760 (p=0.048; medium confidence and small delta AIC) (data not shown). The results from both experiments (the low p-value and substantial improvement relative to the model with one fewer breakpoint - delta AIC) suggest strongly that there was a possible recombination event around position 604-618. Taken these findings, all of the further conducted analysis used the SBP and GARD Inferred Trees as models to overcome the noise introduced by recombination.

Recombination events occur frequently in viruses, both from plants and animals (Lai, 1992). It is well known that sex and recombination lead to the increase of genetic diversity in populations (Grapputo et al., 2005), and that mutation and recombination events are particularly beneficial in the case of genes which are under diversifying selection, e.g. animal virus genes. It is harder to understand what the function is of recombination and sex, when purifying selection dominates. In fact, this is a common pattern in the case of cellular organisms (Castillo-Davis et al., 2004). An explanation for the function of sex and recombination in this case was suggested by Kondrashov (Kondrashov, 1988). His mutational deterministic hypothesis was primarily proposed for evolution of sexual reproduction. According to this hypothesis, mutations are mainly deleterious, which cause mutational loads and create isolates containing many slightly deleterious mutations. On the other hand, recombination creates genotypes with different levels of deleterious mutations. Selection forces remove genotypes with a higher number of deleterious mutations. The Kondrashov hypothesis may explain recombination events discovered in our study. The CP genes from various ASPV strains are under strong purifying selection, which means that mutations are mainly deleterious and are removed by natural selection. There is also at least one recombination event in the studied gene. It might be possible that the function of recombination in such cases is enhancing the speed of removing deleterious mutations by selection.

The following methodologies: SLAC (single-likelihood ancestor counting), FEL (fixed effects likelihood) and REL (random effects likelihood) were used to determine sites under positive selection (Kosakovsky Pond and Frost 2005). All these methods estimate the number and ratio of nonsynonymous and synonymous substitutions for each
site (codon). A consensus scoring approach was applied to determine the sites most probably experiencing positive selection. All three methods found codon 20 to be under positive selection. Additionally codon 111 was identified by SLAC, and codon 340 by FEL - although the dN>dS difference in this case was not significant (Table. 2). The obtained results suggest that there is one position (codon 20) which is under positive selection. Both SLAC and FEL methods used default p-value of 0.1, and REL used Bayes Factor of 40.

The PARRIS (a partitioning approach for robust inference of selection) methodology, described by Scheffler et al., (2006), was used to test if the detected sites (codon 20, 111 and 340) will be robust enough to pass through a maximum likelihood test, designed specifically to avoid misleading recombination events. PARRIS allows synonymous substitution rates to vary across sites as well as tree topologies and branch lengths to change across detected recombination breakpoints. The PARRIS methodology did not detect any positive selection signals at p<0.1 in the compared ASPV CPs.

The codon-based genetic algorithm GABranch method was used to evaluate if the studied isolates evolve due to positive selection. This method can partition automatically all branches of the phylogeny describing non-recombinant data into groups according to dN/dS. The most probable recombination site according to SBP and GARD consensus located between bases 604-618 was used. Therefore, base 612 was chosen as the cutting alignment site to avoid a frameshift. The two resulting alignments were subjected to GABranch algorithm and the results were compared using omega values. As it is shown on Fig. 1, all branches evolve with omega <1, suggesting that purifying selection occurred during the whole analyzed history of ASPV CP. This analysis has indicated that the majority of mutations occurring in analyzed sequences are deleterious. Similar high levels of nucleotide sequence variability as well as low values of \( \frac{dS}{dS} \) have been reported in different genomic regions for other members of the family Betaflexiviridae (Chare and Holmes, 2006; Teycheney et al., 2005), indicating that the attribute is common to the members of this family.

In this paper we showed that ratio of synonymous and nonsynonymous mutations suggests that purifying selection dominates in the case of ASPV. Four different approaches detected strong signal of purifying selection for the majority of codons. Similar results were described for coat protein genes of Grapevine rupestris stem pitting-associated virus (Alabi et al., 2010) and Sweet potato mild mottle virus (SPMMV) (Tugume et al., 2010). In contrast, animal viruses experience strong diversifying selection
which is caused predominantly by pressure from the immune system of the host. Plant viruses are not subjected to such pressure thus purifying selection is the major force shaping their evolution, the coat protein of the ASPV is no exception.

Although purifying selection dominates in the evolution of ASPV capsid protein, three sites having a diversifying selection signal (codon 20, 111 and 340) were detected. This signal may be significant especially in the case of codon number 20, as positive selection was detected using three (SLAC, REL, FEL) out of four methods. Such phenomenon was also observed for SPMMV where 13 amino acids in the P1 proteinase were under positive selection, whereas purifying selection was implicated for the rest of the sites (Tugume et al., 2010). It is not surprising that various codons are under different selection forces. By analogy to animal viruses, it is likely that codon 20 is involved in reaction with host factors. Host resistance may lead to diversifying selection in this position. It is also possible that the amino acid coded by codon 20 is on the surface of capsid protein. Analysis of capsid three-dimensional structure may confirm this hypothesis. Unfortunately there are no proper templates which could be applied for homology modeling of this protein to date. Mutational studies should be conducted to gain more insight about the possible importance of this part of the capsid protein.

References


Tables:

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Accession number</th>
<th>Original</th>
<th>CP gene length [nt]</th>
<th>Position of deletion(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSA-H&lt;sup&gt;a&lt;/sup&gt;</td>
<td>D21828</td>
<td>pear</td>
<td>1245</td>
<td>-</td>
</tr>
<tr>
<td>ST54</td>
<td>AF345892</td>
<td>pear</td>
<td>1191</td>
<td>8006-8059</td>
</tr>
<tr>
<td>ST113</td>
<td>AF345894</td>
<td>pear</td>
<td>1125</td>
<td>8006-8059</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8233-8298</td>
</tr>
<tr>
<td>ST132</td>
<td>AF345894</td>
<td>pear</td>
<td>1125</td>
<td>8006-8059</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8233-8298</td>
</tr>
<tr>
<td>GNKIII/45</td>
<td>AF491929</td>
<td>pear</td>
<td>1185</td>
<td>8487-8493</td>
</tr>
<tr>
<td>GNKVII/34</td>
<td>AF345893</td>
<td>pear</td>
<td>1128</td>
<td>8006-8059</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8233-8295</td>
</tr>
<tr>
<td>PA66&lt;sup&gt;a&lt;/sup&gt;</td>
<td>D21829</td>
<td>apple</td>
<td>1245</td>
<td>-</td>
</tr>
<tr>
<td>MT32</td>
<td>AF438521</td>
<td>apple</td>
<td>1191</td>
<td>8006-8059</td>
</tr>
<tr>
<td>J335</td>
<td>AF491930</td>
<td>apple</td>
<td>1191</td>
<td>8006-8059</td>
</tr>
<tr>
<td>MT24</td>
<td>AF438522</td>
<td>apple</td>
<td>1191</td>
<td>8006-8059</td>
</tr>
<tr>
<td>ST181</td>
<td>AF495382</td>
<td>apple</td>
<td>1191</td>
<td>8006-8059</td>
</tr>
<tr>
<td>N1</td>
<td>AF491931</td>
<td>unknown</td>
<td>1191</td>
<td>8006-8059</td>
</tr>
</tbody>
</table>

Table 1. Characteristics of a coat protein gene of the studied ASPV isolates.

<sup>a</sup>Reference strains kindly provided by W. Jelkmann

<sup>b</sup>Corresponding to the sequence of ASPV PA66 (GenBank accession no. D21829)
Table 2

The results obtained from three methods: SLAC, FEL and REL used to determine possible positive selection in ASPV CP. For each method, two values are present: omega and p-value; with the exception of REL methodology, which uses Bayes Factor instead of p-value. The consensus column, shows which of the methods indicate a certain codon is under positive selection; “P”: positive selection, "P+": statistical significance of such prediction, “N”: negative selection, "N-":statistical significance.

<table>
<thead>
<tr>
<th>Codon</th>
<th>SLAC dN-dS</th>
<th>SLAC p-value</th>
<th>FEL dN-dS</th>
<th>FEL p-value</th>
<th>REL dN-dS</th>
<th>REL Bayes Factor</th>
<th>Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>1.705</td>
<td>0.135</td>
<td>0.826</td>
<td>0.027</td>
<td>0.108</td>
<td>40.416</td>
<td>P, P+, P+</td>
</tr>
<tr>
<td>111</td>
<td>1.222</td>
<td>0.056</td>
<td>-0.651</td>
<td>0.321</td>
<td>-0.533</td>
<td>0.125</td>
<td>P+, N, N</td>
</tr>
<tr>
<td>340</td>
<td>-3.778</td>
<td>1.000</td>
<td>0.104</td>
<td>0.354</td>
<td>-0.182</td>
<td>41.056</td>
<td>N, P, N-</td>
</tr>
</tbody>
</table>
Figure 1. GABranch analysis showing a Neighbor-Joining tree with levels of omega coded by colors. Green branches represent the highest dN/dS values, red and purple colored branches indicate medium to low values, respectively. All the values in this GABranch analysis show omega below 1, suggesting strong purifying selection forces acted during the analyzed ASPV CP history. The tree clearly divides apple and pear virus.
isolates into two separate branches. It might also suggest that isolate of unknown origin
(AF491931) was originally isolated from apple. Also, position of isolates AF345892 and
D21829 in the tree could imply recombination events.

393

394