

Evidence for RNA recombination between distinct isolates of *Pepino mosaic virus*

Beata Hasiów-Jaroszewska^{1,*}, Arnold Kuzniar^{2,*}, Sander A. Peters³, Jack A. M. Leunissen² and Henryk Pospieszny¹

¹Institute of Plant Protection, National Research Institute, Department of Virology and Bacteriology, Poznań, Poland; ²Laboratory of Bioinformatics, Wageningen University and Research Centre (WUR), Wageningen, The Netherlands; ³Bioscience Group, Plant Research International (PRI), Wageningen, The Netherlands

Genetic recombination plays an important role in the evolution of virus genomes. In this study we analyzed publicly available genomic sequences of *Pepino mosaic virus* (PepMV) for recombination events using several bioinformatics tools. The genome-wide analyses not only confirm the presence of previously found recombination events in PepMV but also provide the first evidence for double recombinant origin of the US2 isolate.

Keywords: genetic diversity, PepMV, recombination

Received: 05 March, 2010; revised: 18 June, 2010; accepted: 13 August, 2010; available on-line: 19 August, 2010

INTRODUCTION

A new member of the genus *Potexvirus*, *Pepino mosaic virus* (PepMV), was first described in 1980 (Jones *et al.*, 1980). Since then many genetic variants of PepMV have been isolated from infected plants such as pepino (*Solanum muricatum*) and tomato (*Solanum lycopersicum*) worldwide; however, only a few have been characterized genetically (Ling, 2007; Hanssen *et al.*, 2008; Hasiów *et al.*, 2008). Over the past few years, PepMV has become one of the most important causes of viral diseases in tomato, causing significant economical losses in Europe and North and South America in particular (van der Vlugt *et al.*, 2000; French *et al.*, 2001; Maroon-Lango *et al.*, 2005; Ling, 2007).

PepMV belongs to the group of small RNA viruses (about 6400 nt) whose genomes consist of a single positive-stranded RNA molecule (ssRNA). The PepMV genome encodes five open reading frames (ORFs) in the 5' to 3' order: a putative replicase (ORF1); three movement proteins, TGBp1-3 (ORF2-4); and a coat protein (ORF5) (Maroon-Lango *et al.*, 2005; Ling, 2007; Lopez *et al.*, 2005). The two short (nt 63–85) untranslated regions (UTRs) flanking the coding part of the genome are likely to play a critical role in viral RNA replication and transcription (Dreher, 1999).

The current phylogenetic classification of PepMV isolates distinguishes four genotypes (or strains), namely European (EU), Peruvian (LP), Chilean 2 (Ch2) and Chilean 1 (Ch1/US1) (Ling, 2007; Hasiów *et al.*, 2008). However, the phylogenetic classification of sequences is usually confounded by the presence of recombinant or mosaic sequences (genomes) where distinct sequence regions do not have a common evolutionary descent (Awadalla, 2003). Therefore, recombination events must

be taken into account when drawing conclusions about the evolutionary origins of PepMV isolates.

Recently, RNA recombination has been reported for Belgian PepMV isolates co-infecting tomato plants (Hanssen *et al.*, 2008). In this study two PepMV genotypes, EU and Ch2, were found to recombine in mixed infections under natural conditions. An analysis of a 625-nt fragment of RdRp and TGB1 (nt 3897–4521) revealed that recombinant sequences encode amino-acid sequences that are partially identical to the Ch2 and EU genotypes. In addition, other PepMV genotypes including LP, EU and US have also been found to co-infect tomato plants (Pagán *et al.*, 2006), raising concerns about crop management in the presence of recombinant PepMV variants.

Here we investigate the extent of recombination events as well as reconstruct the most likely evolutionary origins of recombinants in publicly available PepMV sequences using existing bioinformatics tools. The *in silico* analyses not only confirm the presence of recombination events in PepMV as described in the previous study (Hanssen *et al.*, 2008) but, most importantly, provide the first evidence for a double recombinant origin of the US2 isolate, a recombinant between the Chilean 1 and Chilean 2 genotypes.

METHODS

PepMV sequence data. Table 1 lists all PepMV genomic sequences used in this study. The sequences were retrieved from the EMBL/DDBJ/GenBank database using our local Sequence Retrieval System (SRS) (Etzold & Argos, 1993) server (<http://srs.bioinformatics.nl/>).

Pairwise genome comparisons using dot-plots. First, the fully sequenced PepMV genomes were compared with each other in a pairwise manner using a dot-plot analysis. For this, the polydot program (default parameters; EMBOSS package version 6.0.1) was used to show the degree of sequence conservation across each pairwise genome alignment. This analysis enabled genomic regions of distinct evolutionary origins, i.e., putative genome mosaics (or recombinants) to be

✉ e-mail: B.Hasiow@iorpib.poznan.pl

* These authors contributed equally

Abbreviations: bp, base pair; CP, coat protein; DSS, difference of sums of squares; nt, nucleotides; ORFs, open reading frames; PDM, probabilistic divergence measures; PepMV, *Pepino mosaic virus*; RdRp, RNA dependent RNA polymerase; TGB, triple gene block; UTR, untranslated region

inferred by visual inspection of the different pair-wise genome dot-plots.

Statistical inference of mosaic genomes. Potential recombination events between divergent nucleotide sequences were explored with two programs: TOPALi (version 2.0) (Milne *et al.*, 2009) and RECCO (Maydt & Lengauer, 2006). Topali incorporates published recombination detection models, namely a DSS-based one (difference of sums of squares) and the PDM (probabilistic divergence measures) which represent a trade-off between computational efficiency and reliability (Milne *et al.*, 2009). The former model is computationally less demanding than the latter model, but also less reliable in detecting the exact positions of cross-over sites. The following parameters were used to obtain reliable results: significance threshold=0.95, window size=200 bp, and step size=10 bp. The algorithm developed by Maydt and Lengauer (2006) is a fast and sensitive method for detecting recombination events in a set of sequences using a cost minimization approach. As a proof of principle, the previously published Belgian PepMV recombinants (Hanssen *et al.*, 2008) were subjected to the *in silico* analysis whereby the methods could be tested for re-discovering known recombination breakpoints.

Phylogenetic inference in mosaic genomes. Full-length genome sequences of PepMV (Table 1; excluding Belgian isolates) were first aligned using default settings of the MEGA software (version 3) (Kumar *et al.*, 2004). The resulting multiple sequence alignment (MSA) was split into three distinct genomic segments (nt 1–2300, 2301–4600 and 4601–6400) according to the two cross-over sites (denoted as S1 and S2) as indicated by polydot and TOPALi. Each genomic segment was used to infer a phylogenetic tree from the nucleotide sequences using the neighbor-joining algorithm (NJ) as implemented in the MEGA software (Kumar

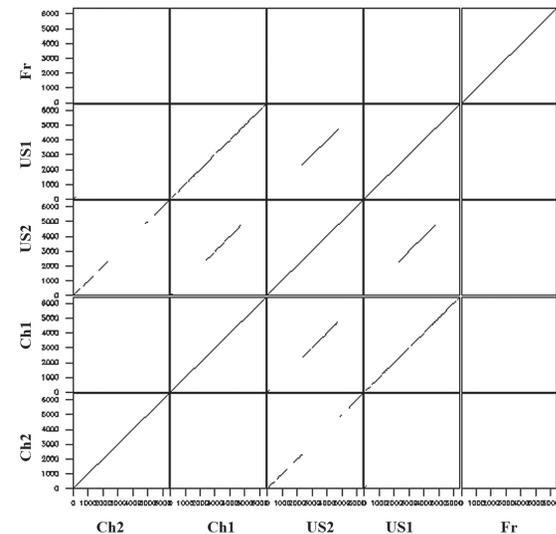


Figure 1. Detection of recombination in PepMV genomes using dot-plot analysis

Pairwise genome comparisons of five PepMV isolates (Fr, US2, Ch1, US1 and US2, accession numbers are given in Table 1) show that US2 is a recombinant between two different genotypes Ch1/US1 and Ch2.

et al., 2003). The reliability of the NJ trees was assessed using 1000 bootstrap replicates.

RESULTS

The whole-genome dot-plots of PepMV isolates provided the first insights into the pairwise similarities and/or differences between the genomes compared (Fig. 1). In particular, the comparisons between US2 and Ch2, US2 and Ch1 (or US1), and Ch1 (or US1) and Ch2 isolates showed clear patterns of sequence conservation/divergence across the genomes: the middle segment of the US2 genome has diverged more than its two flanking regions when compared with the Ch2 isolate, while the pattern of sequence conservation was the opposite when compared with Ch1 or US1. These patterns provide a strong indication that the US2 genome has a recombinant origin, and more specifically, that US2 is a double-recombinant between two evolutionally distinct genotypes, namely Ch1/US1 and Ch2.

Genome-wide scans for expected recombination sites in the two recombinant Belgian PepMV isolates (or mosaics between EU and Ch2 ancestral genotypes) were carried out using two statistical models implemented in the TOPALi program. This 'proof-of-principle' analysis confirmed the previous results on the recombinant origins of the two PepMV isolates, including the genomic positions of the cross-over sites (Hanssen *et al.*, 2008) (Fig. 2). Next, twelve genome sequences were

Table 1. Full-length and partial PepMV genomic sequences used in the study
Accession numbers refer to EMBL/DDBJ/GenBank entries).

Isolate	Accession Number	Genomic sequence	Country of origin
Fr	AJ438767	full	France
Sp-13	AF484251	full	Spain
LE-2000	AJ606359	full	Spain
LE-2002	AJ606360	full	Spain
H	AM491606	full	Hungary
PK	EF408821	full	Poland
Pa	FJ612601	full	Poland
US1	AY509926	full	USA
US2	AY509927	full	USA
Ch1	DQ000984	full	Chile
Ch2	DQ000985	full	Chile
SM.74	AM109896	full	Peru
LP-2001	AJ606361	full	Peru
3206/04/A1	EF599531	partial	Belgium
3206/11/B1_I	EF599546	partial	Belgium
3205/InocB_E	EF599517	partial	Belgium
3206/13/A1_A	EF599534	partial	Belgium

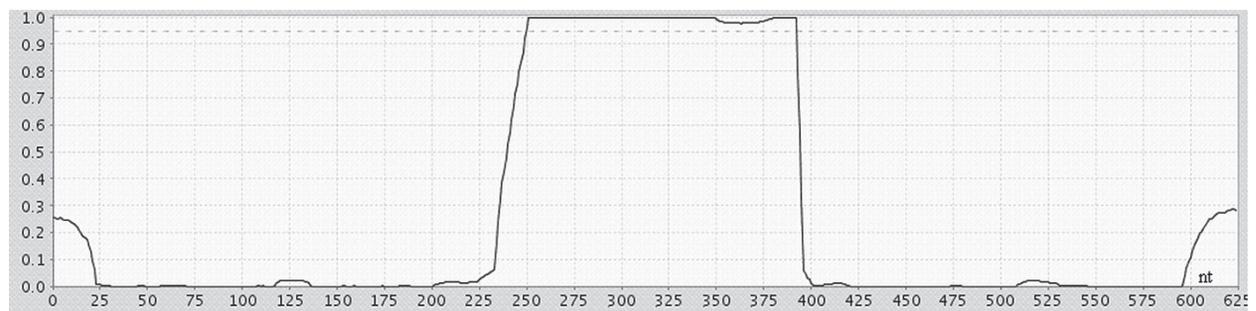


Figure 2. PDM-based method correctly predicts recombination sites in genomes of Belgian PepMV isolates

The prediction is based on multiple genome alignment of two Belgian (EF599546, EF599517), one EU (EF599531) and one Ch2 (EF599534) PepMV isolates.

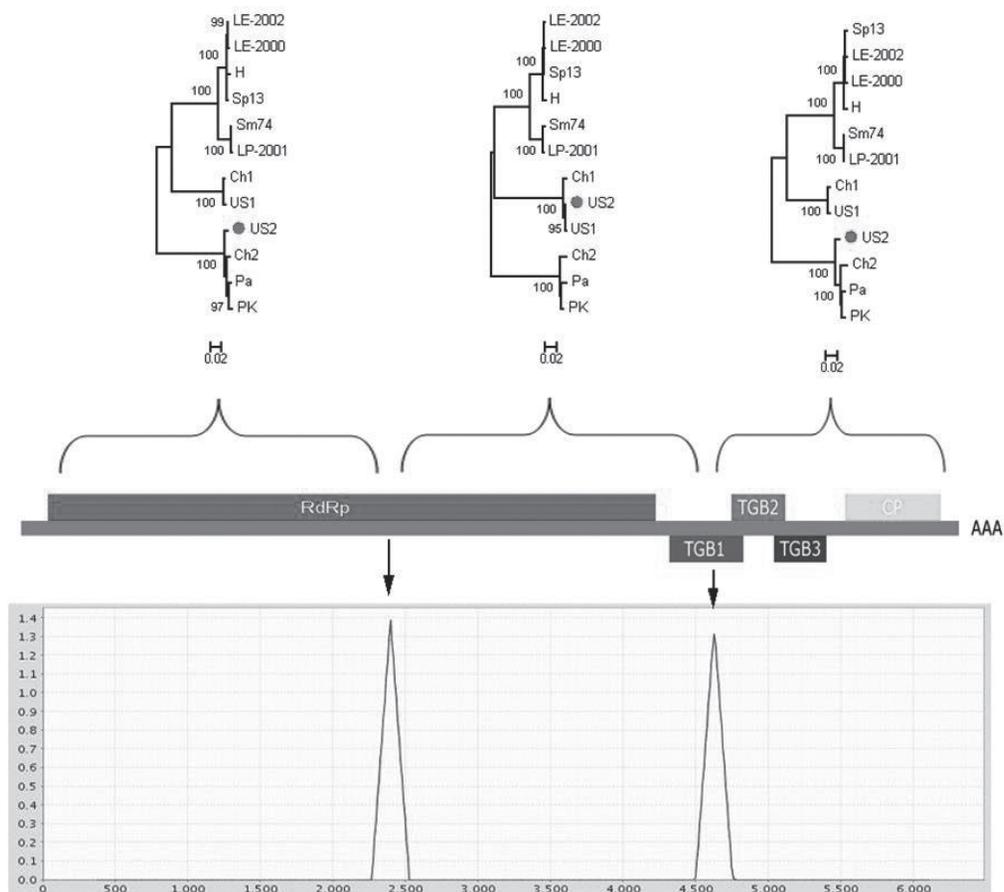


Figure 3. PDM-based method predicts two recombination sites in the US2 genome indicated by arrows

Phylogenetic trees constructed for different coding regions of PepMV genomes show phylogenetic incongruence due to the recombinant US2 isolate.

analyzed for recombination sites using the DSS model. Consequently, two genomic regions of likely recombination were identified: the first one between nucleotides 2077 and 2406, and the second one between 4500 and 4744 (not shown). For more precise delineation of the recombination sites, the PDM model, which is computationally more intensive than the DSS model, was applied only on four out of the twelve PepMV sequences (Fr, Ch1, PK and US2) (Fig. 3). As a result, two significant recombination sites were detected at positions around nt 2300 (in the polymerase gene) and 4700 (in the triple gene block 1) (Fig. 3). No PepMV isolates other than US2 were found to cause the recombination sites to disappear completely when one of the four isolates was repeatedly

removed from the analysis (not shown). Examination of the RECCO program output regarding the occurrence of recombination events between PepMV isolates confirmed that US2 is a recombinant. Three genomic sites of likely recombination were identified: the first one between nucleotides 111 and 121 (corresponding sequence GAAA-GATCAG), second 2310–2311 (corresponding sequence AT), and the third 4744–4747 (corresponding sequence ACTT) dividing the US2 genome into four segments. The first (5'UTR) and third (nt 2310–4744) are homologous to Ch1/US1 sequence whereas the second (nt 122–2310) and fourth (4744–end) are closer to the Ch2 genotype. We also identified an additional recombination site in the 5' UTR (at nt position 111–121) of the US2 isolate using the

RECCO software. Thus, this result supports the hypothesis about the recombinant origin of the US2 genome.

The phylogenetic trees obtained from the three distinct coding parts of the US2 genome provided further evidence for this hypothesis. Specifically, the first and third genomic segments belong to the clade of Ch2 genotypes whereas the second (middle) genomic segment is closer to the Ch1/US1 ancestral genotype (Fig. 3).

DISCUSSION

Recombination has a prominent role in the evolution of RNA viruses because new genetic variants may arise by combining together two or more parental genomes. Several factors such as viral replication proteins, RNA templates and host genes have a bearing on RNA recombination, suggesting complex virus–host interactions. As recombination events usually lead to incorrect or incomplete phylogenies, it is therefore critical to screen for putative recombinants prior to phylogenetic tree reconstruction. The results obtained from independent but complementary genome-wide analyses support the view that recombination events are probably rampant in the PepMV population.

Interestingly, recombination has been detected exclusively in two coding regions, namely the RNA dependent RNA polymerase (RdRp) and triple gene block 1 genes of the PepMV-US2 isolate, where even minor mutations could have a deleterious effect on the function and/or structure of the encoded proteins. Although the number of PepMV recombinants identified experimentally and/or predicted computationally is still very small, we speculate that some genomic regions may be more prone to recombination than others. Such regions, often called “hot spots”, may differ in the primary and/or secondary structure, sequence composition (A-U rich sequences), genomic neighborhood and biological constraints (such as functional or structural), which can promote distinct recombination mechanisms (Lai, 1992; Alejska *et al.*, 2001). In this study we did not address the question of which recombination models provide the most plausible explanation for the recombinant origin of the PepMV-US2 genome. Thus, further research is required to illuminate this subject.

Moreover, the role of recombination in the evolution of PepMV is not well understood. Phylogenetic-based sequence comparison and experimental approaches support the claim that genetic recombination often causes changes in the natural population of plant viruses, resulting in enhanced pathogenicity, extended host range or overcoming host resistance factors (Escriu *et al.*, 2003; Garcia-Arenal *et al.*, 2003). Hanssen *et al.* (2008) observed increased severity of symptoms when tomato plants were infected by the EU and CH2 genotypes. Nonetheless, such effects might equally likely be due to synergistic activities between isolates in natural conditions.

Recently, several necrotic PepMV isolates such as PepMV-Pa, PepMV-Ku and PepMV-Ros have been isolated from greenhouse tomato plants throughout Poland (Hasiów-Jaroszewska *et al.*, 2009). These isolates differ from other Polish isolates in terms of host range and symptoms: they induce severe necrosis on tomato plants and cause local necrotic lesions on thorn-apple (*Datura innoxia*). The phylogenetic analysis based on three distinct genomic regions (the triple gene block 1, coat protein gene, and a part of polymerase gene) support close evolutionary relatedness between the PepMV-Pa necrotic isolate and the Ch2 genotype. Whether these necrotic isolates have recombinant origins, however, remains to be investigated.

These preliminary findings indicate that PepMV can serve as a model for studying host–virus interactions as well as the genotype–phenotype relationships. With this knowledge we are going to undertake *in vivo* and *in vitro* recombination experiments using different PepMV isolates. Moreover, further genomic studies including high-throughput sequencing and computational analyses of new PepMV genomes will be performed to provide valuable insight into the correlation between necrosis and recombination, and ultimately result in better management of plant diseases caused by this virus.

Acknowledgements

This work was supported by grant N N 310 163 438 from the Polish Ministry of Science and Higher Education and by Foundation for Polish Science.

REFERENCE

- Alejska M, Kurzyńska-Kokorniak A, Broda M, Kierzek R, Figlerowicz M (2001) How RNA viruses exchange their genetic material. *Acta Biochim Pol* **48**: 391–407.
- Awadalla P (2003) The evolutionary genomics of pathogen recombination. *Nat Rev Genetic* **4**: 50–60.
- Dreher TW (1999) Functions of the 3'-untranslated regions of positive strand RNA viral genomes. *Annu Rev Phytopathol* **37**: 151–174.
- Escriu E, Fraile A, Garcia-Arenal F (2003) The evolution of virulence in a plant virus. *Evol Int J Org Evol* **57**: 755–765.
- Etzold T, Argos P (1993) SRS — an indexing and retrieval tool for flat file data libraries. *Comput Appl Biosci* **9**: 49–57.
- French CJ, Bouthillier M, Bernardy M, Ferguson G, Sabourin M, Johnson RC, Masters C, Godkin S, Mumford R (2001) First report of *Pepino mosaic virus* in Canada and the United States. *Plant Dis* **85**: 1121.
- Garcia-Arenal F, Fraile A, Malpica JM (2003) Variation and evolution of plant virus populations. *Int Microbiol* **6**: 225–232.
- Hanssen IM, Paelman A, Wittemans L, Goen K, Lievens B, Bragard C, Vanachter C, Thomma BPHJ (2008) Genetic characterization of *Pepino mosaic virus* isolates from Belgian greenhouse tomatoes reveals genetic recombination. *Eur J Plant Pathol* **121**: 131–146.
- Hasiów B, Borodynyo N, Pospieszny H (2008) Complete genomic RNA sequence of the Polish *Pepino mosaic virus* isolate belonging to the US2 strain. *Virus Genes* **36**: 1–8.
- Hasiów-Jaroszewska B, Pospieszny H, Borodynyo N. (2009) New necrotic isolates of *Pepino mosaic virus* representing Ch2 genotype. *J Phytopathol* **157**: 494–496.
- Jones RAC, Koenig R, Lesemann DE (1980) *Pepino Mosaic Virus* a new potyvirus from pepino (*Solanum muricatum*). *Ann Appl Biol* **94**: 61–68.
- Kumar S, Tamura K, Nei M (2004) Mega3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinformatics* **5**: 150–163.
- Lai MM (1992) RNA recombination in animal and plant viruses. *Microbiol Rev* **56**: 61–79.
- Ling KS (2007) Molecular characterization of two *Pepino mosaic virus* variants from imported tomato seed reveals high levels of sequence identity between Chilean and US isolates. *Virus Genes* **34**: 1–8.
- Lopez C, Soler S, Nuez F (2005) Comparison of the complete sequences of three different isolates of *Pepino mosaic virus*: size variability of the TGBp3 protein between tomato and *L. peruvianum* isolates. *Arch Virol* **150**: 619–627.
- Maroon-Lango CJ, Guaragna MA, Jordan RL, Hammond J, Bandla M, Marquardt SK (2005) Two unique US isolates of *Pepino mosaic virus* from a limited source of pooled tomato tissue are distinct from a third (European-like) US isolate. *Arch Virol* **150**: 1187–1201.
- Maydt J, Lengauer T (2006) Recco: recombination analysis using cost optimization. *Bioinformatics* **22**: 1064–1071.
- Milne I, Lindner D, Bayer M, Husmeier D, McGuire G, Marshall DF, Wright F (2009) TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics* **25**: 126–127.
- Pagán I, Cordoba-Selles MC, Martinez-Priego L, Fraile A, Malpica JM, Jorda C, Garcia-Arenal F (2006) Genetic structure of the population of *Pepino mosaic virus* infecting tomato crops in Spain. *Phytopathology* **96**: 274–279.
- van der Vlugt RPA, Stijger CCM, Verhoeven JTJ, Lesemann DE (2000) First report of *Pepino mosaic virus* on tomato. *Plant Dis* **84**: 103.