Sequence characteristics of potato virus Y recombinants

Xiaojun Hu,^{1,2} Alexander V. Karasev,^{1,3} Celeste J. Brown^{2,3} and Jim H. Lorenzen^{2,4}

¹Department of Plant, Soil, and Entomological Sciences, University of Idaho, Moscow, ID 83844, USA

²Bioinformatics and Computational Biology Program, University of Idaho, Moscow, ID 83844, USA ³Department of Biological Sciences, University of Idaho, Moscow, ID 83844, USA

⁴International Institute of Tropical Agriculture, Kampala, Uganda

Potato virus Y (PVY) is one of the most economically important plant pathogens. The PVY genome has a high degree of genetic variability and is also subject to recombination. New recombinants have been reported in many countries since the 1980s, but the origin of these recombinant strains and the physical and evolutionary mechanisms driving their emergence are not clear at the moment. The replicase-mediated template-switching model is considered the most likely mechanism for forming new RNA virus recombinants. Two factors, RNA secondary structure (especially stem-loop structures) and AU-rich regions, have been reported to affect recombination in this model. In this study, we investigated the influence of these two factors on PVY recombination from two perspectives: their distribution along the whole genome and differences between regions flanking the recombination junctions (RJs). Based on their distributions, only a few identified RJs in PVY genomes were located in lower negative FORS-D, i.e. having greater secondary-structure potential and higher AU-content regions, but most RJs had more negative FORS-D values upstream and/or higher AU content downstream. Our wholegenome analyses showed that RNA secondary structures and/or AU-rich regions at some sites may have affected PVY recombination, but in general they were not the main forces driving PVY recombination.

Received 5 June 2009 Accepted 11 August 2009

Correspondence

Alexander V. Karasev

akarasev@uidaho.edu

INTRODUCTION

RNA recombination is a major evolutionary factor in RNA viruses (Roossinck, 1997, 2003; Strauss & Strauss, 1988). The phenomenon has been observed for many plant viruses (reviewed by Simon & Bujarski, 1994), e.g. bromoviruses (Allison *et al.*, 1989), luteoviruses (Gibbs, 1995), nepoviruses (Le Gall *et al.*, 1995a, b) and cucumoviruses (Fraile *et al.*, 1997). It is especially prevalent in potato virus Y (PVY) (Lorenzen *et al.*, 2006, 2008; Revers *et al.*, 1996).

The replicase-mediated template-switching (copy choice) model is considered the most likely mechanism for forming new RNA virus recombinants (Nagy & Simon, 1997). In this model, the viral replicase switches template from the donor to the acceptor RNA during RNA synthesis and uses the nascent RNA as a primer to generate a recombinant RNA (Nagy & Simon, 1997). AU-rich regions could be recombination-promoting signals in viral RNAs (Shapka & Nagy, 2004). They have been found frequently near

A supplementary figure showing a multiple sequence alignment of breakpoint 502 is available with the online version of this paper.

recombination sites in many viruses, e.g. picornaviruses (Pilipenko et al., 1995), brome mosaic virus (BMV) (Nagy & Bujarski, 1996, 1997; Shapka & Nagy, 2004), turnip crinkle virus (TCV) and cucumber necrosis virus (Cheng & Nagy, 2003), citrus tristeza virus (CTV) (Vives et al., 2005), bean pod mottle virus (Zhang et al., 2007), turnip mosaic virus (TuMV) (Ohshima et al., 2007), grapevine fanleaf virus and Arabis mosaic virus (Vigne et al., 2008). On the other hand, RNA secondary structures, especially stemloop structures (hairpins), can promote replicase pausing or act as signals for replicase pausing on the template or nascent RNA (Nagy & Simon, 1997). They are also RNA replication enhancers that may promote RNA recombination directly by binding the replicase-aborted nascentstrand complex during the crossover event (Nagy & Simon, 1997; Nagy et al., 1998). Stem-loop structures were also observed in recombination sites in TCV (Cascone et al., 1993; Nagy et al., 1999b), tombusviruses (White & Morris, 1995) and norovirus (Nayak et al., 2008). However, a comprehensive study of how these two factors affect plant RNA virus recombination based on whole-genome analyses has not yet been conducted.

Π

One method for predicting thermodynamic RNA secondary structure over a viral genome is FORS-D analysis (Forsdyke, 1995; Zhang *et al.*, 2008). A nucleotide segment with a more negative FORS-D value has high potential to develop a stem–loop structure. In human immunodeficiency virus, most recombination breakpoints occurred in more negative FORS-D regions (Zhang *et al.*, 2005). Equivalent analyses have yet to be reported for plant RNA viruses.

PVY is the type member of the genus Potyvirus (family Potyviridae). The PVY genome is a single-stranded, positivesense RNA of about 9.7 kb, with a virus-encoded protein (VPg) attached covalently to its 5' end and a 3' poly(A) tail (Fauquet et al., 2005; Riechmann et al., 1992). The viral RNA encodes a single, large polyprotein, which is processed posttranslationally by three virus-encoded proteases (P1, HC-Pro and NIa) into nine gene products (Fig. 1a) (Dougherty & Carrington, 1988). The major strains, defined based on their pathology in tobacco, are the ordinary strain PVY^O and the necrotic strain PVY^N (de Bokx & Huttinga, 1981). PVY has a high degree of genetic variability (identity range 82-98%, according to a BLAST search) and is also subject to recombination (Revers et al., 1996). Since the early 1980s, a number of PVY recombinants have been documented, including PVY^{NTN}, PVY^N-Wi (in North America termed PVY^{N:O}) (Crosslin *et al.*, 2005; Piche *et al.*, 2004; Singh *et al.*, 2003; Thole et al., 1993). A new recombinant that differs from PVY^{NTN} and PVY^N-Wi, called NE-11, was recently identified (Lorenzen et al., 2008; Piche et al., 2004).

The PVY^N-Wi isolates contain one recombination junction (RJ) between the HC-Pro and P3 region, where the sequence switches from PVY^N-like to PVY^O-like (see Fig. 1a). Some of these isolates also contain an additional RJ in the P1 N-terminal region (Glais *et al.*, 2002). PVY^{NTN} isolates have an N:O:N:O genomic structure with three or four RJs (see Fig. 1a). The first RJ in the HC-Pro/P3 region is nearly identical to that of the PVY^N-Wi RJ; in the 6K2–NIa N-terminal region, the sequence reverts to PVY^N-like sequence; and in the CP C-terminal region, there is another switch to a PVY^O-type sequence (Boonham *et al.*, 2002; Glais *et al.*, 2002; Lorenzen *et al.*, 2006). Other PVY recombination patterns were also described recently (Ogawa *et al.*, 2008).

Although many PVY recombinants have been reported, the mechanisms of selection of quite a limited number of RJs are poorly understood and the physical factors affecting recombination in PVY are not known. In this study, we strive to understand recombination in PVY. First, we identified all RJs among all 43 PVY whole-genome sequences in GenBank (Table 1) and confirmed statistically that Mont (N) and Oz (O) were the most likely parents of N-Wi and NTN recombinants, using the Monte Carlo simulation tool SeqGen and the likelihood recombination-detection software LARD (Holmes *et al.*, 1999). Second, we predicted thermodynamic RNA secondary structure in the parental isolates, Mont (N) and Oz (O), using FORS-D

analysis to investigate the relationship between RNA secondary structures and RJs. Third, we compared the sequences around the breakpoints (or RJs) in the parents with regard to FORS-D values and nucleotide composition. Our whole-genome analyses showed that RNA secondary structures and/or AU-rich regions at some sites affected PVY recombination, but they were not the main forces driving PVY recombination.

METHODS

Recombination analyses. All PVY full-length genome sequences were retrieved from GenBank (http://www.ncbi.nlm.nih.gov/). Retention of only a single sequence from homologues with 100% sequence identity resulted in 43 sequences for subsequent analyses (Table 1). Sequences were aligned by MUSCLE 3.6 using default parameters (Edgar, 2004). Recombination breakpoints were identified by the RDP2 software (Martin et al., 2005), using the aligned nucleotide sequences as input. We assumed PVY^{O} and PVY^{N} to be the potential ancestral (or parental) strains, and PVY^{NTN} and PVY^N-Wi to be potential recombinants. Positions of the identified RJs were verified using several programs implemented in RDP2 with default settings, e.g. RDP (Martin & Rybicki, 2000), MAXCHI (Smith, 1992), GENECONV (Padidam et al., 1999), SiScan (Gibbs et al., 2000) and CHIMAERA (Posada & Crandall, 2001). The criterion for inclusion was a very low empirical P-value ($<10^{-15}$) supported by multiple programs. For accurate detection of crossover sites, only parents and recombinants sharing the same crossover sites were chosen. Alignment files for all of the crossovers are provided in Supplementary Fig. S1 (available in JGV Online).

Mont (\mbox{PVY}^{N}) and $\mbox{Oz}\;(\mbox{PVY}^{O})$ isolates were identified by $\mbox{RDP2}$ as the most likely parents of the N-Wi and NTN strains. For statistical confirmation of crossover events between Mont and Oz at identified RJs, the simulation methods were used as described by Holmes et al. (1999). Briefly, for each RJ of the parents and one recombinant, 1000 nt segments centred around the RJ were retrieved from the alignment and input into PHYML (Guindon & Gascuel, 2003) under the HKY model (Hasegawa et al., 1985). The tree inferred by this method was used to simulate 500 datasets using SeqGen under the HKY model and no recombination. Then, likelihood estimates of the real and simulated data given the tree with and without recombination were determined by LARD (Holmes et al., 1999). The likelihood ratios for the simulated datasets were used to build the null distribution (H₀, no recombination). If the likelihood ratios for the real data fell outside this distribution, the null hypothesis was rejected in favour of the alternative hypothesis (H₁, recombination), providing statistical support for the recombination event.

FORS-D analysis. To investigate the relationship between local stemloop potential and natural recombination, parental isolates Mont (PVY^N) and Oz (PVY^O) were analysed by FORS-D, using a modified Forsdyke method (Forsdyke, 1995). The two full-length sequences were divided into 951 successive 200 nt windows that each overlap the previous window by 190 nt. For each window, 50 randomized sequences were obtained by using the Perl function List :: Util 'shuffle'. The minimum free energy (MFE) values of sequences were calculated by the RNAfold program in the Vienna package (Hofacker *et al.*, 1994). A natural sequence usually has two kinds of information: nucleotide composition and order. By shuffling the order of a sequence, the contribution of the nucleotide order to the stem-loop potential (FORS-D) can be found by subtracting the mean of the MFE values for folding the randomized sequences (FORS-M) from the MFE value for folding the natural sequence (FONS). This provides an estimate of the local base-ordered stem-loop potential.



	Table	1.	Full-length	PVY	sequences	used	in	the	analy	sis
--	-------	----	-------------	-----	-----------	------	----	-----	-------	-----

GenBank accession no.	Isolate	Strain	Original host	Reference	Country
AJ585198	SASA-61	N	Solanum tuberosum	Barker & McGeachy, unpublished data	UK
AY166866	Tu 660	Ν		Nie & Singh (2003)	Canada
AY884984	RRA-1	Ν		Lorenzen et al. (2006)	USA
AY166867	N-Jg	Ν		Nie & Singh (2003)	Canada
AJ890346	Nicola	Ν	Nicotiana tabacum	Schubert et al. (2007)	Germany
DQ157180	NE-11	Ν		Lorenzen et al. (2008)	USA
AJ585197	SCRI-N	Ν	Solanum tuberosum	Barker & McGeachy, unpublished data	UK
AY884983	Mont	Ν		Lorenzen et al. (2006)	USA
X97895	N605	Ν		Jakab et al. (1997)	Switzerland
M95491	Hun-NTN	NTN		Thole <i>et al.</i> (1993)	Hungary
EF026075	PB312	NTN		Baldauf et al. (2006)	USA
AJ890345	Linda	NTN	Nicotiana tabacum	Schubert et al. (2007)	Germany
AJ585342	NIB-NTN	NTN	Solanum tuberosum	Barker et al., unpublished data	Slovenia
AJ890344	Ditta	NTN	Nicotiana tabacum	Schubert et al. (2007)	Poland
AY884982	423-3	NTN		Lorenzen et al. (2006)	USA
AJ890347	Satina	NTN	Solanum tuberosum	Schubert et al. (2007)	Germany
AJ889867	Wilga156	NTN	Solanum tuberosum	Schubert et al. (2007)	Germany
AJ890343	Gr99	NTN	Nicotiana tabacum	Schubert et al. (2007)	Poland
AJ889866	12-94	NTN	Solanum tuberosum	Schubert et al. (2007)	Poland
AJ890342	34/01	NTN	Nicotiana tabacum	Schubert et al. (2007)	Poland
AJ889868	Wilga156var	NTN	Solanum tuberosum	Schubert et al. (2007)	Germany
AB270705	SYR-NB-16	NTN	Nicotiana tabacum	Chikh Ali et al. (2007)	Japan
DQ008213	PN10A	N-Wi		Lorenzen et al. (2006)	USA
AJ890350	Wilga 5	N-Wi	Solanum tuberosum	Schubert et al. (2007)	Germany
AJ890349	LW	N-Wi	Solanum tuberosum	Schubert et al. (2007)	Poland
AJ584851	SASA-207	N-Wi	Solanum tuberosum	Barker & McGeachy, unpublished data	UK
AY745492	L56	N-Wi		Nie et al. (2004)	Canada
AY745491	Mb112	N-Wi		Nie et al. (2004)	Canada
EF026076	PB209	N-Wi		Baldauf et al. (2006)	USA
DQ157178	ID-1	N-Wi		Lorenzen et al. (2008)	USA
AY884985	Alt	N-Wi		Lorenzen et al. (2006)	USA
DQ157179	OR-1	N-Wi		Lorenzen et al. (2008)	USA
X12456	Fr	0		Robaglia et al. (1989)	France
AJ585195	SASA-110	0	Solanum tuberosum	Barker & McGeachy, unpublished data	UK
U09509	O-139	0	Solanum tuberosum	Singh & Singh (1996)	Canada
AJ585196	SCRI-O	0	Solanum tuberosum	Barker & McGeachy, unpublished data	UK
EF026074	Oz	0		Baldauf et al. (2006)	USA
AJ890348	Adgen	С	Solanum tuberosum	Schubert et al. (2007)	France
AF463399	MŇ	Non-pot		Fellers et al. (2002)	
DQ309028	NC57	Non-pot	Nicotiana tobacum	Hari & Kelly, unpublished data	USA
AF237963	NNP	Non-pot	Pepper	Fanigliulo et al. (2005)	Italy
AJ439544	SON41	Non-pot	Solanum nigrum	Moury et al. (2002)	France
AJ439545	LYE84.2	Non-pot	Lycopersicon esculentum	Moury et al. (2002)	Spain

Identification of features related to RJs. The stem-loop potential and compositions of the upstream (5') and downstream (3') sequences of the two parents (Mont and Oz) around RJs were compared. The mean of five FORS-D values before and after the breakpoint represents the value for each flanking side. The statistical significance of differences between two sides was compared by Student's *t*-test. Others have shown that a 25 nt AU-rich sequence can promote recombination (Shapka & Nagy, 2004); therefore, the 25 nt block compositions of the upstream and downstream regions around the breakpoints were compared.

For an overview of the AU distribution along the genome, the AU content was calculated, using sliding windows of size 25 and step 10.

The AU distributions of Mont and Oz were drawn and the RJs were marked.

RESULTS

Recombination-detection analyses

For detecting recombination breakpoints and viewing mosaic structures, we used several recombination-detection programs included in the RDP2 software. Oz and Mont were chosen as the best parents (around 99 % similarity for

2

Pattern	Group	Recombination breakpoint(s)*	Isolates	<i>P</i> -value†
1	N-Wi	2419	SASA-207, L56, Mb112, PB209, ID-1, Alt, OR-1	8.08×10^{-27}
2	N-Wi	502, 2395–2398	LW, Wilga5, PN10A	8.08×10^{-27}
3	N-Wi	652, 2419	261-4	8.08×10^{-27}
4	NTN	2419, 5834, 9193	Linda, Ditta, NIB-NTN, Hun-NTN, PB312, v942490, 423-3	1.48×10^{-18}
5	NTN	2419, 5857, 8607	SYR-NB-16	1.48×10^{-18}
6	NTN	502, 2177, 5834, 9193	Gr99	1.48×10^{-18}
7	NTN	502, 2395–2419, 5834, 9193	PVY-12, 12-94, 34/01	1.48×10^{-18}
8	NTN	502, 2398, 5834, 8585	Wilga156	1.48×10^{-18}
9	NTN	502, 2398, 5834, 6714	Wilga156var	1.48×10^{-18}

Table 2.	Recombination	patterns of PVY	detected by	recombination-detecting	programs
----------	---------------	-----------------	-------------	-------------------------	----------

*Breakpoints indicate positions in the alignment of all sequences. Ambiguous breakpoints are indicated by a range of values.

†Greatest *P*-value among recombinants identified by the recombination-detecting programs RDP, GENECONV, BootScan, MaxChi, CHIMAERA and SiScan in RDP2.

3

the respective daughter sequences) for N-Wi and NTN recombinants, as determined by the lowest *P*-values for the programs RDP ($P < 4.029 \times 10^{-127}$), GENECONV ($P < 2.342 \times 10^{-129}$), BootScan ($P < 7.163 \times 10^{-205}$) and MAXCHI ($P < 2.54 \times 10^{-43}$). Summarizing these recombinants according to the RJs showed nine recombination patterns with eight RJs [Table 2; Fig. 1, panel (iii)]. As mentioned, this did not include all known PVY recombinants, but only those best represented by the postulated parents: Mont (N strain) and Oz (O strain).

To further analyse statistically whether Mont and Oz can form a recombinant at a specific RJ, we used all eight RJs in simulation with Mont, Oz and recombinants. For each RJ, the simulated 500 datasets and the natural data were analysed by LARD, which determined the breakpoint for each dataset by the maximum-likelihood estimate. The null distribution (hypothesis H₀, no recombination) was obtained from simulated sequences' log-likelihood ratios as given by LARD. Its range was 1.3–16.2 for all eight RJs. However, the values for the natural RJs were 54–149 (Table 3), well outside the null distribution, which statistically supported the alternative hypothesis of recombination (H_1) . Therefore, crossover events between Mont and Oz might truly happen in nature. We can use them to analyse the features that affect recombination.

FORS-D and AU distribution along the PVY genome

Previous studies on small RNA viruses have shown that RNA secondary structure and/or AU-rich sequence may promote RNA virus recombination in the templateswitching model (Nagy & Simon, 1997). A stem–loop structure could form a heteroduplex and bring two recombination RNA substrates into close proximity (Nagy & Bujarski, 1993). Thus, both positive-strand and negative-strand RNAs at this region could be the templates for RNA replicase to produce new recombinants. Because FORS-D reflects the local base-ordered stem–loop potential, it can be used to estimate RNA secondary structure in the virus genome. To investigate roles of RNA secondary structure and AU-rich regions in PVY recombination, we

Table 3. Comparison of the log-likelihood ratios (LLRs) for PVY natural recombinant sequences to their respective simulated data at different breakpoints

Breakpoint*	Gene	Recombination patterns	LLRs for natural recombinants†	Range of LLRs for simu- lated datasets
502	P1	2	149.7 ± 4.3	1.6-13.0
652	P1	3	162.5	1.6-12.2
2177	HC-Pro	6	72.5	1.6-14.0
2395-2419	HC-Pro	1-5, 7	77.2 ± 5.4	1.5–16.2
5834-5857	VPg	4–9	89.1 ± 2.8	1.3–14.1
6714	NIa	9	99.8	1.6-13.3
8585-8607	NIb	5, 8	69.3 ± 4.1	1.6-12.3
9193	СР	4, 6, 7	54.4 ± 4.1	1.2–14.8

*Breakpoints indicate positions in the alignment of all sequences. Ambiguous breakpoints are indicated by a range of values. †Mean±sD; for breakpoints 652, 2177 and 6714, only one recombinant was identified, so no sD is given. 5

used the sliding-windows method to compute AU content and FORS-D values along the Mont and Oz genomes and drew their distributions [Fig. 1, panels (ii, iii)].

The AU content along the PVY genome fluctuated [Fig. 1, and the range was 28–84 mol%. AU-rich (≥60 mol% AU) regions in Mont and Oz isolates represented 42 and 37% of the entire genome length, respectively. The most prominent RJs, located at positions 2395–2419 and 5834–5857, were in AU-rich regions (see alignments provided in Supplementary Fig. S1). However, there were no RJs in the PVY genome regions with the highest AU content, e.g. the regions around positions 1445, 3155, 4175, 7645 and 9465, which had AU contents ≥80 mol%.

The distribution of FORS-D values is shown in Fig. 1, panels (iii), in which the eight identified RJs are marked. The overall mean FORS-D value of Mont and Oz isolates of PVY was -1.78 and -2.34 kcal mol⁻¹ from 951 points, respectively; 63 and 69% of Mont and Oz points were negative along the genome, respectively. The mean for the FORS-D values around RJs of Mont and Oz was -3.56 kcal mol⁻¹ in both cases. These values were lower than the means [P=0.04 (statistically significant) for Mont, P=0.11 for Oz]. In a positive-strand FORS-D distribution, RJs 502 and 2177 were located in the more negative FORS-D values for Mont (N parent). At the same time, in a negativestrand FORS-D distribution, RJs 2395-2419 and 5834-5857 were located in the more negative FORS-D values for Oz (O parent). It needs to be pointed out that most of the regions with the lowest FORS-D values had no RJs.

Identification of features related to recombination breakpoints

After analysing the whole genome, we focused attention on the small regions around the RJs to identify sequence features distinguishing identified recombinant breakpoints in the PVY genome. We calculated FORS-D values for the sequences flanking the corresponding RJs of the parents (Mont and Oz) (Table 4). In the positive strand, we found that, in most cases, upstream (5') FORS-D values were lower than downstream (3') values: six of eight in Oz and eight of eight in Mont. This difference was highly significant for four RJs (502, 652, 2177 and 9193) (P<0.01; Table 4) in Mont, and highly significant for one RJ (652) and significant for three RJs (2395-2419, 5834–5857, 9193) (0.01<P<0.05; Table 4) in Oz. We also found a highly significant difference in the FORS-D value for negative-strand RJs, e.g. Mont negative-strand RJ 5834-5857. So, we concluded that the stem-loop structures might enhance RNA recombination at these RJs.

We also examined the AU composition of the sequences on either side of the RJs in Mont and Oz (Table 5). All RJs had at least one AU-rich region, which we defined as having $\geq 60 \mod \%$ AU. RJs 652, 2395–2419 and 5834 had a high AU content for both parents. When we checked regions

Left (5')Right (3')P-value†Left (5')Right (3') <th>Left (5')Right (3')P-value†Left (3')Right (3')<th>Breakpoint*</th><th></th><th>Mont plus</th><th></th><th></th><th>Oz plus</th><th></th><th></th><th>Mont minus</th><th></th><th></th><th>Oz minus</th><th></th></th>	Left (5')Right (3')P-value†Left (3')Right (3') <th>Breakpoint*</th> <th></th> <th>Mont plus</th> <th></th> <th></th> <th>Oz plus</th> <th></th> <th></th> <th>Mont minus</th> <th></th> <th></th> <th>Oz minus</th> <th></th>	Breakpoint*		Mont plus			Oz plus			Mont minus			Oz minus	
502 -15.3 ± 2.1 -8.2 ± 2.0 0.0006^a -7.8 ± 1.0 -7.9 ± 3.2 0.94 -6.8 ± 0.8 -3.1 ± 2.2 0.008^a -1.2 ± 2.0 -5.7 ± 4.2 $0.5.7 \pm 4.2$ 652 -5.1 ± 2.6 2.1 ± 2.0 0.001^a -6.9 ± 2.7 1.4 ± 1.3 0.0002^a -3.5 ± 1.2 -6.9 ± 1.8 0.009^a -1.2 ± 2.6 -3.0 ± 2.6 0.2 2177 -5.4 ± 1.8 1.2 ± 3.5 0.007^a -3.5 ± 1.2 -6.9 ± 1.8 0.009^a -1.2 ± 2.6 -3.0 ± 2.6 0.2 2177 -5.4 ± 1.8 1.2 ± 3.5 0.005^a 2.9 ± 2.1 3.3 ± 5.3 0.87 0.49 ± 0.6 -2.1 ± 2.6 -3.0 ± 2.6 0.2 $2395-2419$ -1.7 ± 1.1 -1.7 ± 2.0 0.96 -3.1 ± 4.5 2.7 ± 2.0 0.03^a -1.6 ± 2.4 0.08^a -1.5 ± 3.7 -1.1 ± 3.1 $5834-5877$ -1.9 ± 2.6 0.996^a -2.7 ± 1.2 1.2 ± 2.8 0.008^a -1.5 ± 2.6 $0.2 \pm 2.2 \pm 3.7$ -1.1 ± 3.1 $5835-8607$ -1.9 ± 2.6 0.99 ± 4.5 0.22 ± 2.0 0.14 <	502 -15.3 ± 2.1 -8.2 ± 2.0 0.006^{d} -7.8 ± 1.0 -7.9 ± 3.2 0.94 -6.8 ± 0.8 -3.1 ± 2.2 0.008^{d} -1.2 ± 2.0 -5.7 ± 4.2 0.06 552 -5.1 ± 2.6 2.1 ± 2.0 0.001^{d} -6.9 ± 2.7 1.4 ± 1.3 0.0002^{d} -3.5 ± 1.2 -6.9 ± 1.8 0.009^{d} -2.1 ± 2.6 -3.0 ± 2.6 0.63 2177 -5.4 ± 1.8 1.2 ± 3.5 0.005^{d} 2.9 ± 2.1 3.3 ± 5.3 0.87 0.49 ± 0.6 -0.9 ± 2.6 0.28 -2.9 ± 0.8 -5.2 ± 3.7 0.001 $2395-2419$ \mp -1.7 ± 1.1 -1.7 ± 2.0 0.96 -3.1 ± 4.5 2.7 ± 2.0 0.03^{b} 1.4 ± 2.0 -1.6 ± 2.4 0.08 -1.5 ± 3.7 -1.1 ± 3.1 0.76 $5834-5857$ \mp -1.9 ± 2.6 -0.8 ± 1.4 0.43 -2.7 ± 1.2 1.2 ± 2.8 0.02^{b} -6.1 ± 2.7 -0.9 ± 1.8 0.008^{d} -0.6 ± 2.0 2.1 ± 2.2 0.07 $5334-5857$ \mp -1.9 ± 2.6 0.9 ± 1.4 0.43 -2.7 ± 1.2 1.2 ± 2.8 0.02^{b} -6.1 ± 2.7 -0.9 ± 1.8 0.008^{d} -0.6 ± 2.0 2.1 ± 2.2 0.07 $5334-5857$ \pm 0.04 ± 1.9 0.99 ± 4.5 0.99 -3.1 ± 4.5 2.7 ± 2.0 0.01^{4} 3.8 ± 1.4 1.5 ± 2.1 0.08 -1.5 ± 3.7 -1.1 ± 2.2 0.07 $5334-5857$ \pm 0.008^{d} -0.6 ± 2.0 2.1 ± 2.2 0.007^{d} -0.5 ± 1.8 -1.0 ± 3.3 0.76 -3.6 ± 1.5 -1.8 ± 2.4 0.19 $2585-8607$ \mp -0.4 ± 1.9 0.99 ± 4.5 0.57 -1.1 ± 1.2 -3.3 ± 3.1 0.17 -1.5 ± 1.8 -1.0 ± 3.3 0.76 -3.6 ± 1.5 -1.8 ± 2.4 0.13 $2585-8607$ \pm -2.9 ± 2.0 3.5 ± 3.7 0.009^{d} -6.8 ± 3.6 -0.5 ± 2.6 0.01^{b} 3.9 ± 1.0 3.6 ± 2.0 0.77 -2.5 ± 1.5 0.09 ± 3.1 0.72 3.5 ± 3.7 0.29 ± 2.0 3.5 ± 3.7 0.009^{d} -6.8 ± 3.6 -0.5 ± 2.6 0.01^{b} 3.9 ± 1.0 3.6 ± 2.0 0.77 -0.7 ± 3.5 0.99 ± 3.1 0.75 * Breakpoints indicate positions in the alignment of all sequences. Ambiguous breakpoints are indicated by a range of values. The true positions for Mont and Oz were used to calculate FO means.		Left (5')	Right (3')	P-value†	Left (5')	Right (3')	<i>P</i> -value†	Left (5')	Right (3')	P-value†	Left (5')	Right (3')	<i>P</i> -value†
$ \begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$	652 -5.1 ± 2.6 2.1 ± 2.0 0.001^a -6.9 ± 2.7 1.4 ± 1.3 0.0002^a -3.5 ± 1.2 -6.9 ± 1.8 0.009^a -2.1 ± 2.6 -3.0 ± 2.6 0.63 2177 -5.4 ± 1.8 1.2 ± 3.5 0.005^a 2.9 ± 2.1 3.3 ± 5.3 0.87 0.49 ± 0.6 -0.9 ± 2.6 0.28 2.9 ± 0.8 -5.2 ± 3.7 0.007 $2395-2419$ -1.7 ± 1.1 -1.7 ± 2.0 0.96 -3.1 ± 4.5 2.7 ± 2.0 0.03^b 1.4 ± 2.0 -1.6 ± 2.4 0.08 -1.5 ± 3.7 -1.1 ± 3.1 0.76 $534-587$ -1.9 ± 2.6 -0.8 ± 1.4 0.43 -2.7 ± 1.2 1.2 ± 2.8 0.02^b -6.1 ± 2.7 -0.9 ± 1.8 0.08^a -1.5 ± 3.7 -1.1 ± 2.1 0.76 574 3.1 ± 2.7 3.2 ± 0.8 0.90 -2.7 ± 1.2 1.2 ± 2.8 0.02^b -6.1 ± 2.7 -0.9 ± 1.8 0.08^a -0.6 ± 2.0 2.1 ± 2.2 0.07 574 3.1 ± 2.7 3.2 ± 0.8 0.90 2.0 ± 1.5 0.2 ± 2.6 0.17 3.8 ± 1.4 1.5 ± 2.1 0.06 2.9 ± 1.7 0.35 $8585-8607$ -0.4 ± 1.9 0.92 -1.1 ± 1.2 -3.3 ± 1.4 0.17 3.6 ± 1.5 -1.8 ± 2.4 0.19 $8585-8607$ -0.4 ± 1.9 0.92 -1.1 ± 1.2 -3.3 ± 1.4 0.17 3.6 ± 1.5 -1.8 ± 2.4 0.19 $8585-8607$ -0.4 ± 1.9 0.92 -1.1 ± 2.2 -3.3 ± 1.4 1.5 ± 2.1 0.02 -2.6 ± 1.5 -1.0 ± 2.4 0.17 $8585-8607$ -0.4 ± 1.9 0.9 ± 4.5 0.009^a -6.8 ± 3.6 $-0.5\pm$	502	-15.3 ± 2.1	-8.2 ± 2.0	0.0006^{a}	-7.8 ± 1.0	-7.9 ± 3.2	0.94	-6.8 ± 0.8	-3.1 ± 2.2	0.008^{a}	-1.2 ± 2.0	-5.7 ± 4.2	0.06
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	652	-5.1 ± 2.6	2.1 ± 2.0	0.001^{a}	-6.9 ± 2.7	1.4 ± 1.3	0.0002^{a}	-3.5 ± 1.2	-6.9 ± 1.8	0.009^{a}	-2.1 ± 2.6	-3.0 ± 2.6	0.63
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	2177	-5.4 ± 1.8	1.2 ± 3.5	0.005^{a}	2.9 ± 2.1	3.3 ± 5.3	0.87	0.49 ± 0.6	-0.9 ± 2.6	0.28	2.9 ± 0.8	-5.2 ± 3.7	0.001^{a}
$\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$	5834-5857‡ -1.9 ± 2.6 -0.8 ± 1.4 0.43 -2.7 ± 1.2 1.2 ± 2.8 0.02^b -6.1 ± 2.7 -0.9 ± 1.8 0.008^a -0.6 ± 2.0 2.1 ± 2.2 0.075 6714 3.1 ± 2.7 3.2 ± 0.8 0.90 2.0 ± 1.5 0.2 ± 2.0 0.14 3.8 ± 1.4 1.5 ± 2.1 0.08^a -0.6 ± 2.6 2.9 ± 1.7 0.35 8585-8607‡ -0.4 ± 1.9 0.9 ± 4.5 0.57 -1.1 ± 1.2 -3.3 ± 3.1 0.17 -1.5 ± 1.8 -1.0 ± 3.3 0.76 -3.6 ± 1.5 -1.8 ± 2.4 0.19 9193 -2.9 ± 2.0 3.5 ± 3.7 0.009^a -6.8 ± 3.6 -0.5 ± 2.6 0.01^b 3.9 ± 1.0 3.6 ± 2.0 0.7 ± 3.5 0.09 ± 3.1 0.72 9193 -2.9 ± 2.0 3.5 ± 3.7 0.009^a -6.8 ± 3.6 -0.5 ± 2.6 0.01^b 3.9 ± 1.0 3.6 ± 2.0 0.7 ± 3.5 0.09 ± 3.1 0.72 * Freedorints indicate positions in the alignment of all sequences. Ambiguous breakpoints are indicated by a range of values. The true positions for Mont and Oz were used to calculate FO	23952419‡	-1.7 ± 1.1	-1.7 ± 2.0	0.96	-3.1 ± 4.5	2.7 ± 2.0	0.03^{b}	1.4 ± 2.0	-1.6 ± 2.4	0.08	-1.5 ± 3.7	-1.1 ± 3.1	0.76
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	6714 3.1 ± 2.7 3.2 ± 0.8 0.90 2.0 ± 1.5 0.2 ± 2.0 0.14 3.8 ± 1.4 1.5 ± 2.1 0.08 1.8 ± 1.9 2.9 ± 1.7 0.35 $8585-8607\ddagger$ -0.4 ± 1.9 0.9 ± 4.5 0.57 -1.1 ± 1.2 -3.3 ± 3.1 0.17 -1.5 ± 1.8 -1.0 ± 3.3 0.76 -3.6 ± 1.5 -1.8 ± 2.4 0.19 9193 -2.9 ± 2.0 3.5 ± 3.7 0.009^a -6.8 ± 3.6 -0.5 ± 2.6 0.01^b 3.9 ± 1.0 3.6 ± 2.0 0.72 -1.8 ± 2.4 0.72 9193 -2.9 ± 2.0 3.5 ± 3.7 0.009^a -6.8 ± 3.6 -0.5 ± 2.6 0.01^b 3.9 ± 1.0 3.6 ± 2.0 0.72 -1.8 ± 2.4 0.72 8585 -2.9 ± 2.0 3.5 ± 3.7 0.009^a -6.8 ± 3.6 -0.5 ± 2.6 0.01^b 3.6 ± 2.0 0.73 -0.7 ± 3.5 0.09 ± 3.1 0.72 8585 -2.9 ± 2.0 3.5 ± 3.7 0.009^a -6.8 ± 2.6 0.01^b 3.9 ± 1.0 3.6 ± 2.0 0.72 -0.7 ± 3.5 0.09 ± 3.1 0.72 8585 0.078 -1.8 ± 2.6 0.01^b	5834-5857‡	-1.9 ± 2.6	-0.8 ± 1.4	0.43	-2.7 ± 1.2	1.2 ± 2.8	0.02^{b}	-6.1 ± 2.7	-0.9 ± 1.8	0.008^{a}	-0.6 ± 2.0	2.1 ± 2.2	0.07
$8585-8607\ddagger -0.4\pm1.9 0.9\pm4.5 0.57 -1.1\pm1.2 -3.3\pm3.1 0.17 -1.5\pm1.8 -1.0\pm3.3 0.76 -3.6\pm1.5 -1.8\pm2.4 0.913 -2.9\pm2.0 3.5\pm3.7 0.009^a -6.8\pm3.6 -0.5\pm2.6 0.01^b 3.9\pm1.0 3.6\pm2.0 0.73 -0.7\pm3.5 0.09\pm3.1 0.9\pm3.1 0.9\pm3.$	8585-8607‡ -0.4 ± 1.9 0.9 ± 4.5 0.57 -1.1 ± 1.2 -3.3 ± 3.1 0.17 -1.5 ± 1.8 -1.0 ± 3.3 0.76 -3.6 ± 1.5 -1.8 ± 2.4 0.19 9193 -2.9 ± 2.0 3.5 ± 3.7 0.009^a -6.8 ± 3.6 -0.5 ± 2.6 0.01^b 3.6 ± 2.0 0.73 -0.7 ± 3.5 0.09 ± 3.1 0.72 *Breakpoints indicate positions in the alignment of all sequences. Ambiguous breakpoints are indicated by a range of values. The true positions for Mont and Oz were used to calculate FO means.	6714	3.1 ± 2.7	3.2 ± 0.8	0.90	2.0 ± 1.5	0.2 ± 2.0	0.14	3.8 ± 1.4	1.5 ± 2.1	0.08	1.8 ± 1.9	2.9 ± 1.7	0.35
9193 -2.9 ± 2.0 3.5 ± 3.7 0.009^a -6.8 ± 3.6 -0.5 ± 2.6 0.01^b 3.9 ± 1.0 3.6 ± 2.0 0.73 -0.7 ± 3.5 0.09 ± 3.1 (-1.7 ± 3.5)	9193 -2.9 ± 2.0 3.5 ± 3.7 0.009^a -6.8 ± 3.6 -0.5 ± 2.6 0.01^b 3.9 ± 1.0 3.6 ± 2.0 0.73 -0.7 ± 3.5 0.09 ± 3.1 0.72 *Breakpoints indicate positions in the alignment of all sequences. Ambiguous breakpoints are indicated by a range of values. The true positions for Mont and Oz were used to calculate FO means.	85858607‡	-0.4 ± 1.9	0.9 ± 4.5	0.57	-1.1 ± 1.2	-3.3 ± 3.1	0.17	-1.5 ± 1.8	-1.0 ± 3.3	0.76	-3.6 ± 1.5	-1.8 ± 2.4	0.19
	*Breakpoints indicate positions in the alignment of all sequences. Ambiguous breakpoints are indicated by a range of values. The true positions for Mont and Oz were used to calculate FO means.	9193	-2.9 ± 2.0	3.5 ± 3.7	0.009^{a}	-6.8 ± 3.6	-0.5 ± 2.6	0.01^{b}	3.9 ± 1.0	3.6 ± 2.0	0.73	-0.7 ± 3.5	0.09 ± 3.1	0.72
		means.												

Ö

Table 4. FORS-D values (kcal mol⁻¹) around breakpoints on both strands of the parent isolates Mont and

#The values flanking these regions were used to calculate the mean FORS-D value

Breakpoint*		Мо	ont			0	Z	
	Left	(5')	Righ	ıt (3')	Left	(5')	Right	t (3')
	A/U	G/C	A/U	G/C	A/U	G/C	A/U	G/C
502	48	52	48	52	52	48	60	40
652	56	44	64	36	60	40	56	44
2177	44	56	60	40	44	56	56	44
2395	64	36	56	44	80	20	60	40
2398	60	40	56	44	80	20	64	36
2419	64	36	36	64	68	32	48	52
5834	72	28	52	48	64	36	56	44
5857	56	44	48	52	60	40	52	48
6714	48	52	64	36	48	52	68	32
8585	56	44	56	44	68	32	56	44
8607	60	40	48	52	60	40	48	52
9193	52	48	60	40	52	48	56	44

Table 5. Sequence composition (mol%) in 25 nt of the detected recombination breakpoints in parent isolates Mont and Oz

*Breakpoints indicate positions in the alignment of all sequences. The true positions for Mont and Oz were used to calculate sequence composition.

more distant from RJs, AU contents were usually around 50 mol%. Hence, we concluded that the AU-rich sequences might play the same role in PVY recombination as in other RNA viruses.

Did both factors affect PVY recombination simultaneously at the RJ? After checking every RJ with two values (FORS-D and AU content), we found that some sites, such as 502, 652, 2177, 2395–2419, 5834–5857 and 9193, had significantly more negative upstream FORS-D values combined with an AU-rich region downstream; some sites, such as 6714 and 8585–8607, only had an AU-rich region. Taken together, we concluded that secondary structure and AUrich regions may affect PVY recombination in some regions.

DISCUSSION

The frequency of viral RNA recombination is affected by several factors, including host genes, the viral replication proteins and various features of the viral RNA templates involved (Nagy, 2008). In this study, we investigated two features of the template: stem-loop structure and AU-rich regions. By analysing the distribution of stem-loop structures and AU-rich regions along the PVY genome, we found that 63 and 69 % of the regions of Mont and Oz had stem–loop potential (FORS-D value <0), respectively, whilst 42 and 37 % of the regions of Mont and Oz were AU-rich ($\geq 60 \mod 8$ AU), respectively. Theoretically, if template features determine locations of crossover sites, many more RJs should be found. However, why are there just a limited number of recombination patterns reported for PVY? There are two possible explanations for this discrepancy. One explanation is that there is a strong selective pressure against successful survival of new PVY

recombinants. In PVY, different genes were reported to be under different selective constraints, e.g. P1 (Ogawa *et al.*, 2008), CP and 6K2 (Moury *et al.*, 2002) were found to evolve by positive selection. Most recombinants formed by the template-switching mechanism thus may not survive. One of the reasons is that imprecise recombination in an AU-rich region may lead to nucleotide additions or deletions. Thus, resulting frameshifts in the PVY genome would result in non-viable recombinants. Another explanation is that other factors, such as host genes, may affect PVY recombination.

When we compared FORS-D value and AU content around RJs, we found that most RJs had the following features: lower negative FORS-D value upstream and/or high AU content downstream, especially in the most common sites (502, 2395-2419, 5834 and 9193), which are usually used for identifying N-Wi and NTN isolates. Many other RNA viruses have AU-rich regions downstream of their recombination sites, but this is the first report of a lower FORS-D value around RJs in a plant virus. Here, we first noticed that upstream FORS-D values are significantly lower than downstream ones at some RJs, such as 502, 652, 2177 and 9193 in Mont and 502, 652 and 9193 in Oz. This means that upstream stem-loop potentials are higher than downstream ones at these RJs. We think that RNA secondary structure might pause or slow down the viral RNA replicase before RJs and promote template switching. However, further biological and biochemical experiments are needed to test this conclusion.

GC-rich (\geq 60 mol% GC content) and AU-rich (\geq 60 mol% AU) regions were termed 'homologousrecombination activators', which can influence homologous recombination (Nagy *et al.*, 1999a). The upstream GCrich region and downstream AU-rich region were observed

X. Hu and others

in RNA recombinants in CTV (Vives *et al.*, 2005), noroviruses (Rohayem *et al.*, 2005) and TuMV (Ohshima *et al.*, 2007). However, in our study, we did not observe higher GC contents upstream of RJs, except for RJ 2419 (data not shown). Thus, GC-rich regions do not seem to play a role in PVY recombination, in contrast to another potyvirus, TuMV (Ohshima *et al.*, 2007). Whether GC-rich regions may have a recombination-silencer function, as in BMV recombination (Nagy & Bujarski, 1998), needs to be determined.

Based on the distribution of FORS-D values and AUcontent variation along the whole PVY genome, we concluded that, whilst RNA secondary structures and AU-rich regions may be enabling forces, they cannot fully explain the recombination observed in PVY. These two factors may play roles in some recombination events; however, other, as-yet-unidentified factors, perhaps host and ecological selection constraints, may drive PVY recombination. One other possible factor affecting PVY recombination is as-yet-unidentified *cis*-acting enhancers of replication (Nagy & Simon, 1997; Nagy *et al.*, 1998), which may also contribute to RJ selection.

ACKNOWLEDGEMENTS

This work was supported in part by USDA-ARS Cooperative Agreements 59-0790-2-066 and 58-1907-8-870 and by the Idaho Potato Commission. We would like to thank Ms Teresa Meacham and Ms Margaret Dibble for technical assistance.

REFERENCES

Allison, R. F., Janda, M. & Ahlquist, P. (1989). Sequence of cowpea chlorotic mottle virus RNAs 2 and 3 and evidence of a recombination event during bromovirus evolution. *Virology* 172, 321–330.

Baldauf, P. M., Gray, S. M. & Perry, K. L. (2006). Biological and serological properties of potato virus Y isolates in Northeastern United States potato. *Plant Dis* **90**, 559–566.

Boonham, N., Walsh, K., Hims, M., Preston, S., North, J. & Barker, I. (2002). Biological and sequence comparisons of potato virus Y isolates associated with potato tuber necrotic ringspot disease. *Plant Pathol* 51, 117–126.

Cascone, P. J., Haydar, T. F. & Simon, A. E. (1993). Sequences and structures required for recombination between virus-associated RNAs. *Science* **260**, 801–805.

Cheng, C.-P. & Nagy, P. D. (2003). Mechanism of RNA recombination in carmo- and tombusviruses: evidence for template switching by the RNA-dependent RNA polymerase *in vitro. J Virol* **77**, 12033–12047.

Chikh Ali, M., Maoka, T. & Natsuaki, K. T. (2007). The occurrence and characterization of new recombinant isolates of PVY displaying shared properties of PVY^NW and PVY^{NTN}. *J Phytopathol* **155**, 409–415.

Crosslin, J. M., Hamm, P. B., Shiel, P. J., Hane, D. C., Brown, C. R. & Berger, P. H. (2005). Serological and molecular detection of tobacco veinal necrosis isolates of *Potato virus Y* (PVY^N) from potatoes grown in the western United States. *Am J Potato Res* **82**, 263–269.

de Bokx, J. A. & Huttinga, H. (editors) (1981). Potato virus Y. In *Descriptions of Plant Viruses*. Kew, UK: Commonwealth Mycological Institute/Association of Applied Biologists.

Dougherty, W. G. & Carrington, J. C. (1988). Expression and function of potyviral gene products. *Annu Rev Phytopathol* **26**, 123–143.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797.

Fanigliulo, A., Comes, S., Pacella, R., Harrach, B., Martin, D. P. & Crescenzi, A. (2005). Characterisation of *Potato virus Y* nnp strain inducing veinal necrosis in pepper: a naturally occurring recombinant strain of PVY. *Arch Virol* 150, 709–720.

Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U. & Ball, L. A. 6 (editors) (2005). Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses. San Diego, CA: Elsevier Academic Press.

Fellers, J. P., Tremblay, D., Handest, M. F. & Lommel, S. A. (2002). The *Potato virus* Y MsNr NIb-replicase is the elicitor of a veinal necrosis-hypersensitive response in root knot nematode resistant tobacco. *Mol Plant Pathol* **3**, 145–152.

Forsdyke, D. R. (1995). A stem-loop "kissing" model for the initiation of recombination and the origin of introns. *Mol Biol Evol* 12, 949–958.

Fraile, A., Alonso-Prados, J. L., Aranda, M. A., Bernal, J. J., Malpica, J. M. & Garcia-Arenal, F. (1997). Genetic exchange by recombination or reassortment is infrequent in natural populations of a tripartite RNA plant virus. *J Virol* 71, 934–940.

Gibbs, M. (1995). The luteovirus supergroup: rampant recombination and persistent partnerships. In *Molecular Basis of Viral Evolution*. Edited by A. J. Gibbs, C. H. Calisher & F. Garcia-Arenal. Cambridge: Cambridge University Press.

Gibbs, M. J., Armstrong, J. S. & Gibbs, A. J. (2000). Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16, 573–582.

Glais, L., Tribodet, M. & Kerlan, C. (2002). Genomic variability in *Potato potyvirus Y* (PVY): evidence that PVY^NW and PVY^{NTN} variants are single to multiple recombinants between PVY^O and PVY^N isolates. *Arch Virol* **147**, 363–378.

Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704.

Hasegawa, M., Kishino, H. & Yano, T. (1985). Dating of the humanape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22, 160–174.

Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M. & Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. *Monatsh Chem* 125, 167–188.

Holmes, E. C., Worobey, M. & Rambaut, A. (1999). Phylogenetic evidence for recombination in dengue virus. *Mol Biol Evol* 16, 405–409.

Jakab, G., Droz, E., Brigneti, G., Baulcombe, D. & Malnoe, P. (1997). Infectious *in vivo* and *in vitro* transcripts from a full-length cDNA clone of PVY-N605, a Swiss necrotic isolate of *Potato virus Y. J Gen Virol* 78, 3141–3145.

Le Gall, O., Candresse, T. & Dunez, J. (1995a). Transfer of the 3' non-translated region of grapevine chrome mosaic virus RNA-1 by recombination to tomato black ring virus RNA-2 in pseudorecombinant isolates. *J Gen Virol* **76**, 1285–1289.

Le Gall, O., Lanneau, M., Candresse, T. & Dunez, J. (1995b). The nucleotide sequence of the RNA-2 of an isolate of the English serotype of tomato black ring virus: RNA recombination in the history of nepoviruses. *J Gen Virol* **76**, 1279–1283.

Lorenzen, J. H., Meacham, T., Berger, P. H., Shiel, P. J., Crosslin, J. M., Hamm, P. B. & Kopp, H. (2006). Whole genome characterization of *Potato virus Y* isolates collected in the western USA and their comparison to isolates from Europe and Canada. *Arch Virol* 151, 1055–1074. Lorenzen, J., Nolte, P., Martin, D., Pasche, J. & Gudmestad, N. (2008). NE-11 represents a new strain variant class of *Potato virus Y*. *Arch Virol* 153, 517–525.

Martin, D. & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16, 562–563.

Martin, D. P., Williamson, C. & Posada, D. (2005). RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* 21, 260–262.

Moury, B., Morel, C., Johansen, E. & Jacquemond, M. (2002). Evidence for diversifying selection in *Potato virus Y* and in the coat protein of other potyviruses. *J Gen Virol* **83**, 2563–2573.

Nagy, P. D. (2008). Recombination in plant viruses. In *Plant Virus Evolution*, pp. 133–156. Edited by M. J. Roossinck. Heidelberg: Springer.

Nagy, P. D. & Bujarski, J. J. (1993). Targeting the site of RNA–RNA recombination in brome mosaic virus with antisense sequences. *Proc Natl Acad Sci U S A* 90, 6390–6394.

Nagy, P. D. & Bujarski, J. J. (1996). Homologous RNA recombination in brome mosaic virus: AU-rich sequences decrease the accuracy of crossovers. J Virol 70, 415–426.

Nagy, P. D. & Bujarski, J. J. (1997). Engineering of homologous recombination hotspots with AU-rich sequences in brome mosaic virus. J Virol 71, 3799–3810.

Nagy, P. D. & Bujarski, J. J. (1998). Silencing homologous RNA recombination hot spots with GC-rich sequences in brome mosaic virus. *J Virol* 72, 1122–1130.

Nagy, P. D. & Simon, A. E. (1997). New insights into the mechanisms of RNA recombination. *Virology* 235, 1–9.

Nagy, P. D., Zhang, C. & Simon, A. E. (1998). Dissecting RNA recombination *in vitro*: role of RNA sequences and the viral replicase. *EMBO J* 17, 2392–2403.

Nagy, P. D., Ogiela, C. & Bujarski, J. J. (1999a). Mapping sequences active in homologous RNA recombination in brome mosaic virus: prediction of recombination hot spots. *Virology* 254, 92–104.

Nagy, P. D., Pogany, J. & Simon, A. E. (1999b). RNA elements required for RNA recombination function as replication enhancers *in vitro* and *in vivo* in a plus-strand RNA virus. *EMBO J* 18, 5653–5665.

Nayak, M. K., Balasubramanian, G., Sahoo, G. C., Bhattacharya, R., Vinje, J., Kobayashi, N., Sarkar, M. C., Bhattacharya, M. K. & Krishnan, T. (2008). Detection of a novel intergenogroup recombinant Norovirus from Kolkata, India. *Virology* **377**, 117–123.

Nie, X. & Singh, R. P. (2003). Evolution of North American PVYNTN strain Tu 660 by mutation rather than recombination. *Virus Genes* 26, 39–47.

Nie, X., Singh, R. P. & Singh, M. (2004). Molecular and pathological characterization of N:O isolates of the *Potato virus Y* from Manitoba, Canada. *Can J Plant Pathol* **26**, 573–583.

Ogawa, T., Tomitaka, Y., Nakagawa, A. & Ohshima, K. (2008). Genetic structure of a population of *Potato virus Y* inducing potato tuber necrotic ringspot disease in Japan; comparison with North American and European populations. *Virus Res* **131**, 199–212.

Ohshima, K., Tomitaka, Y., Wood, J. T., Minematsu, Y., Kajiyama, H., Tomimura, K. & Gibbs, A. J. (2007). Patterns of recombination in turnip mosaic virus genomic sequences indicate hotspots of recombination. *J Gen Virol* 88, 298–315.

Padidam, M., Sawyer, S. & Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology* 265, 218– 225.

Piche, L. M., Singh, R. P., Nie, X. & Gudmestad, N. C. (2004). Diversity among *Potato virus Y* isolates obtained from potatoes grown in the United States. *Phytopathology* **94**, 1368–1375. Pilipenko, E. V., Gmyl, A. P. & Agol, V. I. (1995). A model for rearrangements in RNA genomes. *Nucleic Acids Res* 23, 1870–1875.

Posada, D. & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A* **98**, 13757–13762.

Revers, F., Le Gall, O., Candresse, T., Le Romancer, M. & Dunez, J. (1996). Frequent occurrence of recombinant potyvirus isolates. *J Gen Virol* 77, 1953–1965.

Riechmann, J. L., Lain, S. & Garcia, J. A. (1992). Highlights and prospects of potyvirus molecular biology. *J Gen Virol* 73, 1–16.

Robaglia, C., Durand-Tardif, M., Tronchet, M., Boudazin, G., Astier-Manifacier, S. & Casse-Delbart, F. (1989). Nucleotide sequence of potato virus Y (N strain) genomic RNA. *J Gen Virol* **70**, 935–947.

Rohayem, J., Munch, J. & Rethwilm, A. (2005). Evidence of recombination in the norovirus capsid gene. J Virol 79, 4977–4990.

Roossinck, M. J. (1997). Mechanisms of plant virus evolution. Annu Rev Phytopathol 35, 191.

Roossinck, M. J. (2003). Plant RNA virus evolution. Curr Opin Microbiol 6, 406–409.

Schubert, J., Fomitcheva, V. & Sztangret-Wisniewska, J. (2007). Differentiation of *Potato virus Y* strains using improved sets of diagnostic PCR-primers. *J Virol Methods* 140, 66–74.

Shapka, N. & Nagy, P. D. (2004). The AU-rich RNA recombination hot spot sequence of *Brome mosaic virus* is functional in tombusviruses: implications for the mechanism of RNA recombination. *J Virol* 78, 2288–2300.

Simon, A. E. & Bujarski, J. J. (1994). RNA–RNA recombination and evolution in virus-infected plants. *Annu Rev Phytopathol* 32, 337–362.

Singh, M. & Singh, R. P. (1996). Nucleotide sequence and genome organization of a Canadian isolate of the common strain of *Potato virus Y* (PVY^O). *Can J Plant Pathol* **18**, 209–214.

Singh, R. P., McLaren, D. L., Nie, X. & Singh, M. (2003). Possible escape of a recombinant isolate of *Potato virus* Y by serological indexing and methods of its detection. *Plant Dis* 87, 679–685.

Smith, J. M. (1992). Analyzing the mosaic structure of genes. J Mol Evol 34, 126–129.

Strauss, J. H. & Strauss, E. G. (1988). Evolution of RNA Viruses. Annu Rev Microbiol 42, 657–683.

Thole, V., Dalmay, T., Burgyán, J. & Balázs, E. (1993). Cloning and sequencing of potato virus Y (Hungarian isolate) genomic RNA. *Gene* 123, 149–156.

Vigne, E., Marmonier, A. & Fuchs, M. (2008). Multiple interspecies recombination events within RNA2 of *Grapevine fanleaf virus* and *Arabis mosaic virus*. Arch Virol 153, 1771–1776.

Vives, M. C., Rubio, L., Sambade, A., Mirkov, T. E., Moreno, P. & Guerri, J. (2005). Evidence of multiple recombination events between two RNA sequence variants within a *Citrus tristeza virus* isolate. *Virology* **331**, 232–237.

White, K. A. & Morris, T. J. (1995). RNA determinants of junction site selection in RNA virus recombinants and defective interfering RNAs. *RNA* 1, 1029–1040.

Zhang, C.-Y., Wei, J.-F. & He, S.-H. (2005). The key role for local base order in the generation of multiple forms of China HIV-1 B'/C intersubtype recombinants. *BMC Evol Biol* 5, 53.

Zhang, C., Gu, H. & Ghabrial, S. A. (2007). Molecular characterization of naturally occurring RNA1 recombinants of the *Comovirus bean pod mottle virus*. *Phytopathology* **97**, 1255–1262.

Zhang, C.-Y., Wei, J.-F., Wu, J.-S., Xu, W.-R., Sun, X. & He, S.-H. (2008). Evaluation of FORS-D analysis: a comparison with the statistically significant stem–loop potential. *Biochem Genet* **46**, 29–40.

Dear Authors,

Please find enclosed a proof of your article for checking.

When reading through your proof, please check carefully authors' names, scientific data, data in tables, any mathematics and the accuracy of references. Please do not make any unnecessary changes at this stage. All necessary corrections should be marked on the proof at the place where the correction is to be made; please write the correction clearly in the margin (if in the text they may be overlooked).

Any queries that have arisen during preparation of your paper for publication are listed below and indicated on the proof. Please provide your answers when returning your proof.

Please return your p	proof by Fax (-	+44 (0)118 988	1834) within 2 d	ays of receipt.
----------------------	-----------------	----------------	------------------	-----------------

Query no.	Query
1	Author: do you wish to cite a specific chapter of the book by Fauquet et al.? (See also query 6)
2	Author: 'average' has been replaced here by 'mean'. Is this change correct?
3	Author: please provide a reference for BootScan (this program was not listed in the Methods).
4	Author: 'average' has been replaced here by 'mean'. Is this change correct?
5	Author: 'average' has been replaced three times in this paragraph by 'mean'. Are these changes correct?
6	Author: do you wish to cite a more specific chapter of this book? (See also query 1)
7	Author: 'average' has been replaced by 'mean' in two of the footnotes to Table 4. Are these changes correct?

101/	J
	0
JUV	$_{}$ λ

lournal of General /irology

Offprint Order Form

PAPER	vir014142	Please quote this number in any correspondence
Authors	X. Hu and others	Date
I would l	ke 25 free offprints, plus	additional offprints, giving a total of offprints
Dispatch	address for offprints (BL	OCK CAPITALS please)

Please complete this form **even if you do not want extra offprints.** Do not delay returning your proofs by waiting for a purchase order for your offprints: the offprint order form can be sent separately.

Please pay by credit card or cheque with your order if possible. Alternatively, we can invoice you. All remittances should be made payable to 'Society for General Microbiology' and crossed 'A/C Payee only'.

Tick one

- □ Charge my credit card account (give card details below)
- □ I enclose a cheque/draft payable to Society for General Microbiology
- □ Purchase order enclosed

Return this form to: JGV Editorial Office, Marlborough House, Basingstoke Road, Spencers Wood, Reading RG7 1AG, UK.

		CHARG	ES FO	r addi	TIONAL	OFFPF	RINTS			
Copies	25	50	75	100	125	150	175	200	Per 25 extra	
No. of pages										OFFICE USE ONL
1-2	£23	£40	£58	£76	£92	£110	£128	£145	£23	Issue:
3-4	£35	£58	£81	£104	£128	£150	£173	£191	£29	Vol/part:
5-8	£46	£76	£104	£133	£162	£191	£219	£249	£35	Page nos:
9-16	£58	£92	£128	£162	£196	£231	£267	£301	£40	Extent:
17-24	£70	£110	£151	£191	£231	£272	£312	£353	£46	Price:
each 8pp extra	£18	£23	£29	£35	£40	£46	£53	£58		Invoice: IR/

PAYMENT BY CREDIT CARD (Note: we cannot accept American Express)

 Please charge the sum of £______ to my credit card account.

 My Mastercard/Visa number is (circle appropriate card; no others acceptable):

 Expiry

 date

 Signature:

 Date:

 Cardholder's name and address*:

*Address to which your credit card statement is sent. Your offprints will be sent to the address shown at the top of the form.