dN/dS-based methods detect positive selection linked to trade-offs between different fitness traits in the coat protein of Potato virus Y

Benoît Moury, Vincent Simon

Version définitive du manuscrit publié dans / Final version of the manuscript published in : Molecular Biology and Evolution, 2011, 28 (9), 2707-2717 DOI: 10.1093/molbev/msr105. The publisher version is available at http://mbe.oxfordjournals.org/content/28/9/2707.full.pdf+html.

dN/dS-based methods detect positive selection linked to trade-offs between different fitness traits in the coat protein of *Potato virus Y*

Intented as Research Article

Benoît Moury and Vincent Simon
Institution and current affiliation: INRA, UR407 Pathologie Végétale, F-84140 Montfavet, France

Corresponding author: Benoît Moury
Address : INRA, Unité de Pathologie Végétale, BP94, F-84140 Montfavet, France.
Tel: +33(0)4 32 72 28 16. Fax : +33(0)4 32 72 28 42.
E-mail : Benoit.Moury@avignon.inra.fr

Key words: positive selection, potyvirus, vector, aphid, fitness, trade-off, dN/dS

Running head: Positive selection and trade-offs in potato virus Y

List of nonstandard abbreviations: CP: coat protein; LRT: likelihood ratio test; dN/dS: ratio between the nonsynonymous and synonymous substitution rates; ML: maximum likelihood; NJ: neighbour joining.

Abstract

The dN/dS ratio between nonsynonymous and synonymous substitution rates has been used extensively to identify codon positions involved in adaptive processes. However, the accuracy of this approach has been questioned and very few studies have attempted to validate experimentally its predictions. Using the coat protein (CP) of *Potato virus Y* (PVY; genus *Potyvirus*, family Potyviridae) as a case study, we identified several candidate positively-selected codon positions which differed between clades. In the CP of the N clade of PVY, positive selection was detected at codon positions 25 and 68 by both the softwares PAML and HyPhy. We introduced nonsynonymous substitutions at these positions in an infectious cDNA clone of PVY and measured the effect of these mutations on virus accumulation in its two major cultivated hosts, tobacco and potato, and on its efficiency of transmission from plant to plant by aphid vectors. The mutation at codon position 25 significantly modified the virus accumulation in the two hosts while the mutation at codon position 68 significantly modified the virus accumulation in one of its hosts and its transmissibility by aphids. Both mutations were involved in adaptive trade-offs. We suggest that our study was particularly favourable to the detection of adaptive mutations using dN/dS estimates because, as obligate parasites, viruses undergo a continuous and dynamic interaction with their hosts which favours the recurrent selection of adaptive mutations and because trade-offs between different fitness traits impede (or at least slow down) the fixation of these mutations and maintain polymorphism within populations.

Introduction

Since the beginning of the years 2000, methods based on estimation of the ratio dN/dS between the nonsynonymous (dN) and the synonymous (dS) substitution rates in an alignment of amino-acid-coding sequences (named "dN/dS methods" in the rest of the article) have been used extensively to identify individual codon positions evolving under positive selection (Nielsen 2005). However, both the theoretical grounds of these methods and their accuracy to detect adaptive codon substitutions have been matters of concern (for example Suzuki and Nei 2004; Wyckoff et al. 2005; Galtier and Duret 2007; Hughes 2007, 2008; Kryazhimskiy and Plotkin 2008; Nozawa et al. 2009). Although highly desirable, experimental validation (or refutation) of these predicted positively-selected codon positions is presently extremely scarce. In their study of vertebrate vision genes, Yokoyama et al. (2008) surprisingly found no correspondence between functionally-important amino acid replacements during vertebrate evolution and amino acid positions under positive selection according to dN/dS methods.

Here, we used dN/dS methods to detect candidate amino acid positions for positive selection in a viral protein and we validated these predictions by measuring the effect of nonsynonymous substitutions at these positions on the virus fitness. For this, we focussed on *Potato virus Y* (PVY) (genus *Potyvirus*, family Potyviridae), a major pathogen of crops belonging to the family Solanaceae, including potato, tobacco and pepper. PVY genome consists of a single-stranded positive-sense RNA which encodes a polyprotein further matured into eleven functional proteins. A previous study suggested that 12% of amino acid positions in the PVY coat protein (CP) were evolving under slight positive selection (Moury et al. 2002). However, the small number of sequences available at that time did not allow a more precise clade-by-clade analysis. More than 100 PVY CP sequences are now available, making a thorough analysis of the selection intensity in this protein feasible.

PVY is transmitted horizontally from plant to plant by more than 40 aphid species (family Aphididae) in a nonpersistent manner (Edwardson and Christie 1997) where the virus is retained only transiently by the vector. Vertical transmission of PVY from infected plants to next generation through tubers is also a major problem for the production of commercial seed potatoes. The CP of potyviruses is involved in indirect binding of the virions to the aphid mouthparts (Peng et al. 1998; Blanc et al. 1997) and in plant infection (more specifically in virus translocation within plants; Dolja et al. 1994, 1995; Andersen and Johansen 1998; Atreya et al. 1995; López-Moya and Pirone 1998) and can therefore be a target of selection by both host plants and vectors.

Materials and Methods

Data collection and phylogenetic analyses

Sequences of the PVY CP coding region were downloaded from GenBank in December 2008. Redundant sequences and sequences containing nucleotide ambiguities were discarded from the dataset. Sequences were aligned using the CLUSTALW program (Thompson et al., 1994). The sequence alignment was analysed with RDP version 2 software (Martin et al., 2005) implementing several algorithms to detect recombinant sequences.

Phylogenetic methods were used to determine the branching pattern of PVY sequences and to perform analysis of the selection intensity along the CP. The maximum likelihood (ML) method implemented in PhyML version 3.0 (Guindon and Gascuel 2003), incorporating the best-fit nucleotide substitution model estimated by MODELTEST (Posada and Crandall, 1998), and the neighbour joining (NJ) method implemented in MEGA version 4 software (Tamura et al. 2007) were used to infer the tree topology. Bootstrap resampling was performed to assess branch support. To identify codon sites with positive selection, *i.e.* sites showing a dN/dS ratio significantly higher than one (Kimura, 1983), ML approaches implemented in the HyPhy (Pond and Frost, 2005a) and in the PAML version 4.2 (Yang 2007) packages were performed using the majority-rule consensus tree topology obtained with the NJ method. In HyPhy, the SLAC (Single-Likelihood Ancestor Counting), FEL (fixed effects likelihood) and REL (random effects likelihood) methods (Pond and Frost 2005a, 2005b; Pond et al. 2006) were used. Under PAML, the nested evolution models M0,

M1a, M2a and M3 from one part, and the nested models M7 and M8 from the other part were compared using likelihood ratio tests (LRTs) to assess the significance of positive selection in the sequence alignments and empirical Bayes methods allowed identifying individual positively-selected codon sites *a posteriori* (Yang et al. 2000; Yang 2007). Each PAML model was run at least three times with different initial values for dN/dS to avoid local maximum likelihood estimates.
The LRT developed by Knudsen and Miyamoto (2001) was used to detect specific nucleotide sites that evolve at different rates in the different PVY clades. A significant rate difference between two subgroups at a given site would, thereby, mean that the function of this position could be different in the two clades and/or that evolutionary constraints differ between PVY clades. The likelihood of the null hypothesis assuming that a given position evolves with different rates in the two sequence groups is compared with the likelihood of the alternative hypothesis (same rate in the two subgroups). The program is available at http://www.daimi.au.dk/~compbio/rateshift/ and allows analysis of 30 sequences of a maximum length of 600 nucleotides at the same time. Consequently, two random subsets of 30 sequences encompassing nucleotide positions 1 to 600 or 202 to 801 were analyzed for each clade.

Mutagenesis of an infectious PVY clone

An infectious cDNA clone of PVY isolate N605, a non-recombinant isolate of clade N, was built previously (Jakab et al. 1997; Bukovinszki et al. 2007). This clone was modified to include a cassette containing the 2µ yeast replication origin and a selectable marker (*Trp-1* promoter and gene) as described in Fernandez-Delmond et al. (2004) and Ayme et al. (2006), allowing homologous recombination in yeast. A cDNA clone corresponding to the CP-coding region of N605 was mutated at nucleotide position 74 ($A_{74}T$, *i.e.* adenosine to thymidine substitution changing the asparagine at position 25 of the CP to an isoleucine) or at nucleotide position 202 ($G_{202}A$, *i.e.* guanosine to adenosine substitution changing the glutamic acid at position 68 of the CP to a lysine) using the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA). The region coding the CP was deleted by reverse PCR from the N605 clone and a unique *Not*I restriction site was introduced at the junction, yielding the acceptor shuttle vector. Yeast was cotransformed with the CP-deleted vector linearized with *Not*I and with cDNA fragments of the CP coding region containing the nucleotide substitution which yielded through homologous recombination the mutated N605 clones N605-CP25 and N605-CP68. The conformity of the clones was checked by sequencing to rule out spurious mutations. Primary inoculations with the cDNA clones were made by DNA-coated tungsten particle bombardments of *Nicotiana clevelandii* juvenile plants (4 weeks post sowing), because direct bombardment of potato or tobacco plants was not efficient.

Comparison of accumulation of PVY N605 wild-type clone and mutants in tobacco and potato plants

The fitness of the N605-CP25 or N605-CP68 mutant clones was compared to that of the parental N605 clone in both single-inoculation and competition experiments in tobacco (*Nicotiana tabacum* cv. Xanthi) and potato (*Solanum tuberosum* cv. Bintje). Inocula were calibrated using DAS-ELISA as shown by Ayme et al. (2006) and used separately for single inoculations or mixed for competition experiments.
For single inoculations, the relative accumulation of each variant was measured at 30 days post inoculation (dpi) in plants inoculated either by N605, N605-CP25 or N605-CP68. For each plant, all leaves were pooled, ground with buffer (1:4 wt/vol), and fivefold successive dilutions in buffer of each plant extract were tested by DAS-ELISA as in Ayme et al. (2006). The relative PVY accumulation in these samples was determined using absorbance values measured at 405 nm (A_{405}) where the curves representing A_{405} in function of the sample dilution factor were linear and parallel between samples (Ayme et al. 2006). Since the two amino acid substitutions introduced in the CP of the N605 clone could modify their reactivity towards the polyclonal antibodies used for DAS-ELISA and could consequently introduce a bias in our estimation of the relative concentration of the different PVY variants in plants, we measured the reactivity of these variants as above, using

DAS-ELISA with several dilutions of virus of known absolute concentration purified from Xanthi plants at 10 to 14 dpi and quantified (Moury et al. 2007). Comparison of DAS-ELISA reactivity of the PVY variants was done with the results obtained from four independent virus purifications with a Kruskal-Wallis nonparametric test. A Kruskal-Wallis test was also performed on the relative accumulation of each PVY variant, with and without correction for relative DAS-ELISA reactivity, in order to detect significant accumulation differences between the PVY variants.

For competition experiments, inocula were mixed at a 1:1 ratio based on DAS-ELISA and used to inoculate plants. Calibration of the relative concentration of inocula using this method was shown to be similar to a calibration using a quantitative RT-PCR method (Janzac et al. 2010). The competitiveness of the PVY variants was analyzed using dCAPS analysis (Michaels and Amasino 1998; Neff et al. 1998) at 30 dpi. For each inoculated plant, total RNAs were extracted from pools of three infected leaves with the Tri Reagent kit (Molecular Research Center Inc., Cincinnati, OH) and used to amplify part of the CP coding region by RT-PCR with primers 'dCAPS-N25I' and 'Reverse' or 'dCAPS-E68K' and 'Reverse' (Supplementary Table S1), depending on the competition. The presence of the mutations $A_{74}T$ and $G_{202}A$ in the virus populations was assessed with the endonucleases *Eco*RI and *Mse*I, respectively, which cleave RT-PCR products obtained with N605-CP25 or N605-CP68, but not with N605. Plants were classified into three categories depending on the observation of cleaved or uncleaved RT-PCR products, or both, after 2% agarose gel electrophoresis. Fisher exact tests were performed to compare the competitiveness of the virus variants and to analyse differential selection by the two plant species. For this, the mixed-infected plants were shared between the sets of plants infected by the wild-type or the mutant virus only. Then, to compare the virus competitiveness, the observed number of plants infected by the wild-type or the mutant virus was compared to the numbers expected if the variants had equal fitnesses, *i.e*. to the 1:1 distribution corresponding to the inoculum ratio. Differential selection by host species was assessed for each virus competition by comparing the observed numbers of plants infected by the wild-type or the mutant virus at 30 dpi in tobacco and potato.

The sequence of the CP coding region of the progeny of the cloned PVY mutants in 10 randomly-chosen inoculated tobacco and potato plants was determined to exclude that second-site mutations in this region had modified the fitness properties of the virus or had hindered dCAPS analysis.

The infectivity of N605 and the two derived CP mutants was evaluated by manual inoculation of a series of diluted virus solutions prepared from calibrated inocula to 20 Xanthi plants per dilution and analysis of the frequency of infected plants as assessed by DAS-ELISA.

Comparison of the transmission efficiency of PVY N605 clone and mutants by aphids

Wingless aphids of the species *Myzus persicae* Sulzer or *Aphis gossypii* Glover were starved during 3 to 4 hours prior acquisition. Sets of ten aphids were allowed an acquisition access period of 2 min ± 30 sec on PVY-infected leaves of the tobacco cultivar Xanthi. Each tobacco leaf was used as virus source for a single set of ten aphids. Aphids were then transferred with a paintbrush to healthy Xanthi seedlings at the two-expanded-leaf stage (one aphid per test plant). Aphids were maintained on plants overnight and treated with an insecticide before being transferred to an insect-proof greenhouse for one month. The infection status of inoculated plants was then checked by symptom reading and DAS-ELISA on apical, uninoculated leaves. The relative concentration of PVY in the Xanthi leaves that were used for virus acquisition by aphids was determined by DAS-ELISA as described above.

Results

The pattern of selection in the CP-coding region of PVY varies across clades

After discarding redundant sequences and sequences containing ambiguous nucleotides, a dataset of 125 sequences of the CP coding region (801-nucleotide long) was obtained and aligned. No gaps were found in the alignment. Recombination in the CP-coding region of PVY is well documented (Revers et al. 1996 Ogawa et al. 2008) and datasets containing recombinant sequences can provide false positives in selection analyses (Anisimova et al. 2003; Shriner et al. 2003). Our purpose was to measure the selection intensity in the PVY CP in a clade-by-clade manner. Consequently, we

discarded clades that contained less than 10 sequences, since a reliable estimation of selection intensity was not possible with so few sequences. This provided a dataset composed of 119 sequences (Supplementary Material) for which the phylogenetic trees generated with PhyML or MEGA were consistent, had high bootstrap support at major branches and separated PVY isolates among four clades (Fig. 1), including three nonrecombinant clades (named clades C, N and O; Moury et al. 2010) and a clade of recombinant isolates showing a recombination breakpoint between nucleotide positions 603 and 616 (Revers et al. 1996) of the CP coding region (named "recombinant clade" in the following). Members of this recombinant clade show close similarity with isolates of clades N and O for the 5' and 3' parts of the CP coding region, respectively. Recombination was validated with RDP and by separate phylogenetic analyses of the two CP coding regions. When this recombinant clade was withdrawn, no more recombination signal was detected with RDP. The selection intensity was analysed separately for the four different clades and clade N was analysed together with the recombinant clade for the region spanning nucleotides 1 to 600, upstream of the recombination breakpoint (Table 1).

Using PAML, no codon site under positive selection was detected for clade C, while significant positive selection was detected for amino acid positions 10, 11 and 26 for clade O and for amino acid positions 25 and 68 in the alignment of the 600 nucleotides at the 5' extremity of the CP coding region of the set of isolates belonging to clade N or to the recombinant clade (Table 1). The dN/dS ratio of the positively-selected category of codons was estimated to be 5.6 for the N (and recombinant) clade and 6.7 for clade O (Table 1). These candidate codon positions for positive selection were confirmed with the REL method implemented in HyPhy (Table 1). This method detected also codon positions 26 and 139 for the N (and recombinant) clade and codon positions 9, 15, 16, 24, 30 and 138 for clade O. The SLAC and FEL methods detected a fewer number of codon sites under positive selection for clade O and no codon sites for the N (and recombinant) clade (Table 1). None of the three methods detected codon sites under positive selection for clade C.

In addition, clade N and the recombinant clade were analysed separately for the whole CP coding region, using both PAML and HyPhy. These supplementary analyses did not provide additional candidate codon sites submitted to positive selection. On the contrary, the previously-identified codon sites were detected with lower statistical support, which was probably due to the fact that splitting the sequence dataset in two subsets decreased the statistical power of analyses.

The fact that different amino acid positions in the CP, depending on the clades, appeared to be under positive selection suggests that selective pressures and/or substitution rates at those particular positions could vary between clades. To confirm this possibility, a LRT (Knudsen and Miyamoto 2001) was used to detect evolution rate shifts between the different PVY clades at the codon positions previously detected as candidates for positive selection with both PAML and HyPhy. Codon position 25 was shown to evolve significantly more rapidly in clades C, N or in the recombinant clade than in clade O (Table 2). Also, codon position 26 evolved faster in clades O or N than in clade C, while codon position 68 evolved faster in the recombinant clade than in clades C or O. Finally, codon position 10 evolved faster in clade O than in clade C. No significant rate shift was observed between clades for codon position 11, which was detected as positively selected in clade O (Table 1). Overall, these results add support to most of the codon positions candidate for positive selection and indicate that positive selection at individual codon positions is specific for one or several PVY clades. Application of Knudsen and Miyamoto's (2001) LRT to the other nucleotide positions of the CP coding region did not reveal more positions showing evolutionary rate shifts than under random (at the 5% type-I error threshold).To validate the prediction that the previously-identified codon positions are evolving under positive selection, we measured the effect of nonsynonymous substitutions at these positions on the virus fitness. For this purpose, PVY clones differing only by these mutations should be compared. No cDNA infectious clones corresponding to PVY isolates from clade O or from the recombinant clade are available. Consequently, we used the cDNA clone from isolate N605 (Jakab et al. 1997; Bukovinszki et al. 2007), belonging to clade N, and we selected the codon positions detected as candidates for positive selection in clade N and the recombinant clade both by HyPhy and PAML, *i.e.* codon positions 25 and 68 (Table 1). At position 25, 55 isolates of 62 from clade N or the recombinant clade, including

N605, possessed an asparagine (Asn) codon, four isolates possessed an isoleucine (Ile) codon, while three isolates possessed a serine or a lysine codon. At position 68, 50 isolates of 62, including N605, possessed a glutamic acid codon (Glu), eleven isolates possessed a lysine (Lys) codon, while one isolate possessed a valine codon. Only the effects of the most frequent amino acid substitution observed in clade N and in the recombinant clade at these two positions were examined, *i.e.* effects of the $Asn_{25}Ile$ and $Glu_{68}Lys$ codon substitutions.

Effect of mutations at positively-selected codon positions on PVY fitness in different host plants

We compared the relative accumulation of N605 and its mutants N605-CP25 and N605-CP68, which differ from N605 only by the $Asn_{25}Ile$ or by the $Glu_{68}Lys$ codon substitutions in the CP coding region, in tobacco and potato, the two major cultivated hosts of PVY isolates from clade N (Kerlan and Moury 2008; Moury 2010). Two different inoculation protocols were used. In competition experiments, N605 and either N605-CP25 or N605-CP68, were co-inoculated manually to the plants using inocula composed of similar concentrations of the PVY variants. In single inoculation experiments each virus variant was inoculated manually to a separate set of plants after calibration of the concentration of the virus solutions. In these latter experimental conditions, virus accumulations were compared by DAS-ELISA (Janzac et al. 2010) and it was therefore important to examine the influence of the $Asn_{25}Ile$ and $Glu_{68}Lys$ substitutions in the CP of N605 on the reactivity of the different PVY variants toward the polyclonal antibodies used in DAS-ELISA. Using purified virus of known absolute concentration we did not observe any significant differences of DAS-ELISA reactivity among the three PVY variants (Kruskal-Wallis test, p-value=0.35). N605-CP25 and N605-CP68 showed average DAS-ELISA reactivities 14% and 18% lower than N605, respectively. However, four independent virus purification experiments could be insufficient to detect a true difference in DAS-ELISA reactivity between PVY variants. Consequently, the ELISA data were analysed both with and without correction by the relative reactivity factor between variants.

In the Xanthi cultivar of tobacco, the N605-CP25 mutant accumulated at a significantly higher rate than N605, both in competition situation (p-value=0.008; Fisher exact test) or after single inoculation (Fig. 2). In this host, there was no accumulation difference between N605 and the N605-CP68 mutant whatever the inoculation procedure (with or without competition between the two variants) (Fig. 2).

In the Bintje cultivar of potato and in the single inoculation procedure, N605-CP68 accumulated at a much higher rate than N605 (about 10 times more), while there was no accumulation difference between N605 and N605-CP25 (Fig. 2). In this host, results from competition experiments differed from results obtained after single inoculation: N605 was more competitive than either the N605-CP25 or the N605-CP68 mutant (p-value=0.04 and p-value<0.0001, respectively; Fisher exact tests; Fig. 2). Importantly, the relative competitiveness of each mutant against N605 was significantly different in the two host plants (p-value<0.0001 whatever the plant species; Fisher exact tests).

In addition, there was no significant difference of infectivity in Xanthi plants among the three PVY variants N605, N605-CP25 and N605-CP68, when measuring the proportion of infected plants after manual inoculation with serially-diluted solutions of calibrated virus concentration (data not shown).

Effect of mutations at positively-selected codon positions on PVY transmission efficiency by aphids

PVY, as the other members of the genus *Potyvirus*, can be transmitted by a number of aphid species in a nonpersistent manner, requiring only very short acquisition and inoculation access periods. Using leaves of tobacco cv. Xanthi plants infected by either N605 or its CP mutants as inoculum sources, we performed transmission experiments to healthy Xanthi plants with *Myzus persicae*, one of the most efficient and prevalent PVY vectors. In four independent experiments, a larger number of plants were infected after inoculation by N605-CP68 than after inoculation by N605 (Table 3). In contrast, there was no significant difference between the number of plants infected after inoculation

by N605-CP25 and N605. An experiment performed with *Aphis gossypii*, another aphid species, confirmed that more plants were infected after inoculation by N605-CP68 than by N605 (Table 3). The relative virus concentration in the Xanthi leaves used for virus acquisition by aphids was controlled *a posteriori* by DAS-ELISA after each experiment. For each independent experiment and for all leaves used as inoculum sources, the concentration of N605-CP25 was higher than or equal to that of N605 (data not shown). In contrast, the concentration of N605-CP68 was lower than or equal to that of N605 (data not shown). Since the virus concentration in the source plants and the susceptibility of inoculated plants were not higher for N605-CP68 than for N605, the higher frequency of plants infected by N605-CP68 could be due only to a higher aphid transmission efficiency of this mutant compared to N605. This difference in transmission efficiency by aphids could in turn be the result of a difference in the binding affinity of the virus to the aphid mouth parts where it is retained after acquisition or of a difference of virus release efficiency from the binding sites during the inoculation phase. Indeed, Blanc et al. (1997) showed a positive correlation between aphid transmission efficiency of *Tobacco vein mottling virus* (TVMV; genus *Potyvirus*) and the binding affinity between its CP and the helper component-proteinase (HC-Pro), the other viral protein required for aphid transmission. Alternatively, a difference in virus distribution in cells and tissues of source plants or a virus-induced difference in the probing behaviour of the vectors could affect the accessibility of the virus variants to the vectors and consequently their transmission efficiency.

Discussion

Identifying mutations responsible for adaptation

This study constitutes one of the first experimental validations of predicted positively-selected codon positions by measuring the impact on virus fitness of nonsynonymous substitutions at these positions. Indeed, substitutions at two amino acid positions of PVY CP detected by PAML and HyPhy as positively selected among isolates from clade N or from an NxO recombinant clade modified two fitness traits of significant epidemiological relevance: within-plant virus accumulation and between-plants transmission efficiency. The $Asn_{25}Ile$ mutation significantly increased PVY accumulation and competitiveness in tobacco but decreased its competitiveness in potato (Fig. 2). The $Glu_{68}Lys$ mutation significantly increased PVY accumulation in absence of competition in potato, but strongly decreased its competitiveness in that host species (Fig. 2). In addition, the $Glu_{68}Lys$ mutation roughly doubled the virus transmission rate from tobacco to tobacco by aphids (Table 3).

Obviously, the amino acid positions detected as positively selected in the CP of PVY are not randomly distributed. Overall, four clusters of adjacent positions contain 10 of the 12 positions detected by the HyPhy software: positions 9 to 11, 15 and 16, 24 to 26 and 138 and 139 (Table 1). This suggests that particular domains of the CP and the functions that they determine are involved in virus adaptation and targeted by selection. We did not validate the amino acid positions candidate for positive selection that were detected for isolates belonging to clade O of PVY because of the lack of an infectious cDNA clone representing this clade. However, although there is presently no model of the three-dimensional structure of potyvirus particles at the atomic level (Kendall et al. 2008), knowledge of the functional domains of the CP of potyviruses can help us raise hypotheses about the selective forces driving a rapid evolution at these positions. An aspartic acid-alanine-glycine (DAG) amino acid triplet, corresponding to positions 6 to 8 of the N-terminal part of the PVY CP, was shown to be highly conserved among potyviruses and essential in their transmission by aphids (Atreya et al. 1991, 1995) by being involved in binding to the HC-Pro (Blanc et al. 1997; Seo et al. 2010). This region of the CP of potyviruses was also shown to be exposed at the surface of the virions. A remarkable cluster of amino acid positions adjacent or close to the DAG motif was detected as under positive selection with HyPhy, involving positions 9 to 11 and 15 and 16 (Table 1). Positions 10 and 11 were also detected by PAML (Table 1). It was shown that mutations at amino acid position 8 of the CP of TVMV (Atreya et al. 1995), which aligns with amino acid position 9 of the CP of PVY, modified substantially the transmissibility of the virus by aphids. We therefore suspect that amino acid substitutions at some of these five positions could also be

involved in variation of PVY aphid transmissibility. The region spanning amino acid positions 133 to 148 of the CP of *Soybean mosaic virus* (SMV; genus *Potyvirus*) (corresponding to positions 136 to 151 of the CP of PVY) was also shown to be involved in binding to the HC-Pro (Seo et al. 2010). Consequently, positive selection at positions 138 and 139 of the CP of PVY (Table 1) could also be driven by aphid transmission efficiency. Although the point mutations introduced at some of these positions by Seo et al. (2010) did not modify the binding affinity of the CP to HC-Pro, it cannot be excluded that these mutations could modify the transmission efficiency of SMV by aphids. Moreover, it remains possible that other kinds of substitutions than those tested by Seo et al. (2010) or substitutions at adjacent positions could alter aphid transmission efficiency of potyviruses. Concerning the putative positively-selected amino acid positions 24, 26 and 30 (Table 1), no particular functional hypotheses can be formulated, except from their proximity with position 25 which was shown to affect differentially viral accumulation in host plants but not aphid transmission efficiency.

Vectors can impose strong positive selection on plant virus populations

While the intensity of selection exerted by plant hosts on viruses has been estimated in a number of studies (Jenner et al. 2002; Moury et al. 2004; Schirmer et al. 2005; Janzac et al. 2010; Fraile et al. 2011) and a large number of mutations that confer adaptation to particular plant species or genotypes has been identified (Liang et al. 2002; Agudelo-Romero et al. 2008, Meshi et al. 1988; Ayme et al. 2006; Pinel-Galzi et al. 2007), our knowledge of the selection exerted by biological vectors on plant virus populations is comparatively much more limited and arises from two main experimental approaches. First, repeated manual inoculations allowed the selection of virus mutants that lost or decreased their transmissibility by vectors. These viruses carry mutations with strong effects in genes that affect, directly or indirectly, their binding to vector mouthparts or vector cellular receptors (Perry et al. 1998; Andrejeva et al. 1999; Blanc et al. 1997, 1998; Ng and Perry 1999; Ng et al. 2000; Sin et al. 2005). Second, directed mutagenesis of viral proteins known to interact with vectors allowed identifying mutations that affect transmission efficiency (Atreya et al. 1995; López-Moya et al. 1999; Flasinski and Cassidy 1998; Seo et al. 2010). The mutations identified by these methods do not provide an accurate image of virus evolution in natural conditions. Indeed, compared to natural selection, the phenotypic effect of these mutations isexcessively large and negatively skewed. It is noticeable that these experimental approaches did not identify amino acid position 68 (or the surrounding region) of the CP of potyviruses as being important for aphid transmission, probably because it is not predicted to be exposed outside the virus particle and does not affect directly the binding of the CP to the HC-Pro. Our detection of amino acid positions involved in transmission efficiency by vectors in virus genomes differs fundamentally from these previous studies. It allowed identifying a mutation which is the target of positive selection by aphid vectors in field PVY isolates and provided an estimate of the strength of selection exerted on this mutation.

Detection of positive selection and evolutionary trade-offs

The present study emphasizes the importance of evolutionary trade-offs in plant virus evolution. Indeed, mutations at amino acid positions 25 and 68 of the PVY CP display different kinds of antagonistic pleiotropic effects on viral fitness traits. The first kind of trade-off, illustrated by mutation $Glu_{68}Lys$, arises during interaction of the virus with two different organisms: its hosts and vectors. That mutation increased aphid transmission efficiency but decreased drastically the virus competitiveness in potato plants. The second kind of trade-off involved differential virus competitiveness in its hosts, as mutation $Asn_{25}Ile$ which increased PVY competitiveness in tobacco but decreased its competitiveness in potato. Third, a trade-off occurs when a mutation increases or decreases the relative accumulation of the virus in a host depending on the presence or absence of competitors in the viral population, as observed for mutation $Glu_{68}Lys$ in potato which was advantageous in single inoculation and deleterious when the mutant was in competition with the wild-type virus. Finally, a fourth kind of trade-off would have been due to differential virus transmission efficiency by different vector species or genotypes. This kind of trade-off was not

observed in our study when we compared the transmission efficiency of N605 and the N605-CP68 mutant by *M. persicae* and *A. gossypii*, probably because of indirect interaction between the vectors and the CP. By contrast, *Cucumber mosaic virus* (CMV; genus *Cucumovirus*; family Bromoviridae) shows amino acid substitutions in the CP that are involved in trade-offs for transmission by the two aphid species *M. persicae* and *A. gossypii* (Perry et al. 1998) and several of these amino acid positions were detected by PAML as targets of positive selection in natural populations (Moury 2004). This is certainly due to the fact that, contrary to potyviruses, the CP of CMV interacts directly with the aphid mouth parts. By maintaining polymorphism at adaptive codon positions within virus populations, these trade-offs favour the detection of positive selection.

Why does selection intensity vary across PVY clades?

Identifying evolutionary pressures that lead to new adaptations is more challenging and speculative than identifying causal mutations because past environments that exerted these pressures can be different from present ones and these pressures are not unidirectional but rather the result of complex networks of forces in ecosystems as well as in the organisms' genomes.

A first kind of evolutionary pressure is related to internal constraints within viral genomes. Since the different viral genes and proteins should work co-ordinately at the structural and functional levels, recombination events within virus genomes could create sub-optimal combinations of genome regions and could cause an accelerated replacement of (positively-selected) amino acids in some proteins. This mechanism could explain why positive selection was detected in the CP of PVY isolates from clades N and O or from the recombinant clade but not in the CP of clade C isolates. Indeed, a majority of PVY isolates from clades N and O that we have defined according to the CP coding region possess recombination breakpoints elsewhere in their genome, while only a minority of isolates from clade C were shown to be recombinants (Schubert et al. 2007; Hu et al. 2009a, 2009b; Ogawa et al. 2008). However, this mechanism does not explain why the positively-selected amino acid positions are involved in trade-offs between fitness traits.

An alternative explanation involves the relaxation of external constraints on virus evolution during the virus epidemiological cycle. Infections by PVY isolates from clades N and O (and N×O recombinant groups) are mostly observed in potato or tobacco crops while clade C PVY isolates are mostly observed in other solanaceous crops such as pepper (Moury 2010). In pepper or tobacco, plant-to-plant transmission of PVY occurs exclusively by aphids while in potato an additional route for PVY infection is vertical transmission from mother plants to progeny tubers. Consequently, for clades N and O, but not for clade C, the selective constraint of being aphid transmissible can be relaxed over several plant generations by planting PVY-contaminated seed potatoes. In turn, relaxed evolutionary constraints could allow the emergence of positively-selected mutations advantageous for adaptation in potato host plants even if they are deleterious for aphid transmission, as is mutation $Lys_{68}Glu$ in the CP. This alternative mechanism would explain both the differences in positive selection observed between clade C and the other clades of PVY isolates and also the fact that these positively-selected positions are involved in trade-offs between fitness traits.

Acknowledgments

We thank Thierry Candresse, Hervé Lecoq and Philippe Roumagnac for their helpful comments on an earlier version of the manuscript, Catherine Rys, Patrick Gognalons and Nathalie Boissot for raising the aphids and the greenhouse team for plant care.

Literature cited

Agudelo-Romero P, de la Iglesia F, Elena SF. 2008. The pleiotropic cost of host-specialization in tobacco etch potyvirus. Inf Genet Evol. 8:806-814.

Andersen K, Johansen IE. 1998. A single conserved amino acid in the coat protein gene of pea seed-borne mosaic potyvirus modulates the ability of the virus to move systemically in Chenopodium quinoa. Virology 241:304-311.

Andrejeva J, Puurand U, Merits A, Rabenstein F, Jarvekulg L, Valkonen JPT. 1999. Potyvirus helper component-proteinase and coat protein (CP) have coordinated functions in virus-host interactions and the same CP motif affects virus transmission and accumulation. J Gen Virol. 80:1133-1139.

Anisimova M, Nielsen R, Yang Z. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164:1229–1236.

Atreya PL, Atreya CD, Pirone TP. 1991. Amino acid substitutions in the coat protein result in loss of insect transmissibility of a plant virus. Proc Natl Acad Sci U S A. 88:7887-7891.

Atreya PL, López-Moya JJ, Chu M, Atreya CD, Pirone TP. 1995. Mutational analysis of the coat protein N-terminal amino acids involved in potyvirus transmission by aphids. J Gen Virol. 76:265-270.

Ayme V, Souche S, Caranta C, Jacquemond M, Chadœuf J, Palloix A, Moury B. 2006. Different mutations in the genome-linked protein VPg of Potato virus Y confer virulence on the pvr23 resistance in pepper. Mol Plant-Microbe Interact. 19:557-563.

Blanc S, López-Moya JJ, Wang R, García-Lampasona S, Thornbury DW, Pirone TP. 1997. A specific interaction between coat protein and helper component correlates with aphid transmission of a potyvirus. Virology 231:141-147.

Blanc S, Ammar ED, Garcia-Lampasona S, Dolja VV, Llave C, Baker J, Pirone TP. 1998. Mutations in the potyvirus helper component protein: effects on interactions with virions and aphid stylets. J Gen Virol. 79:3119-3122.

Bukovinszki A, Götz R, Johansen E, Balázs E. 2007. The role of the coat protein region in symptom formation on Physalis floridana varies between PVY strains. Virus Res. 127, 122-125.

Dolja VV, Haldeman R, Robertson NL, Dougherty WG, Carrington JC. 1994. Distinct functions of capsid protein in assembly and movement of etch potyvirus in plants. EMBO J. 13:1482-1491.

Dolja VV, Haldeman-Cahill R, Montgomery AE, Vandenbosch KA, Carrington JC. 1995. Capsid protein determinants involved in cell-to-cell and long distance movement of tobacco etch potyvirus. Virology 206:1007-1016.

Edwardson JR, Christie RG. 1997. Potyviruses. In: Florida Agricultural Experiment Station Monograph Series 18-II – Viruses Infecting Pepper and Other Solanaceous Crops. Gainesville, FL: University of Florida.

Fernandez-Delmond I, Pierrugues O, De Wispelaere M, Guilbaud L, Gaubert S, Diveki Z, Godon C, Tepfer M, Jacquemond M. 2004. A novel strategy for creating recombinant infectious RNA virus genomes. J Virol Methods 121:247-257.

Flasinski S, Cassidy BG. 1998. Potyvirus aphid transmission requires helper component and homologous coat protein for maximal efficiency. Arch Virol. 143:2159-2172.

Fraile A, Pagán I, Anastasio G, Sáez E, García-Arenal F. 2011. Rapid genetic diversification and high fitness penalties associated with pathogenicity evolution in a plant virus. Mol Biol Evol. 28:1425-1437..

Galtier N, Duret L. 2007. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. Tr Genet. 23:273-277.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52:696-704.

Hu XX, He CZ, Xiao Y, Xiong XY, Nie XZ. 2009a. Molecular characterization and detection of recombinant isolates of potato virus Y from China. Arch Virol. 154:1303-1312.

Hu XJ, Karasev AV, Brown CJ, Lorenzen JH. 2009b. Sequence characteristics of Potato virus Y recombinants. J Gen Virol. 90:3033-3041.

Hughes AL. 2007. Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. Heredity 99:364–373.
Hughes AL. 2008. The origin of adaptive phenotypes. Proc Natl Acad Sci U S A. 105:13193–13194.
Jakab G, Droz E, Brigneti G, Baulcombe D, Malnoë 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.
Janzac B, Montarry J, Palloix A, Navaud O, Moury B. 2010. A point mutation in the polymerase of Potato virus Y confers virulence toward the Pvr4 resistance of pepper and a high competitiveness cost in susceptible cultivar. Mol Plant-Microbe Interact. 23:823-830.
Jenner CE, Wang X, Ponz F, Walsh JA. 2002. A fitness cost for Turnip mosaic virus to overcome host resistance. Virus Res. 86:1-6.
Kendall A, McDonald M, Bian W, et al. (14 co-authors). 2008. Structure of flexible filamentous plant viruses. J Virol. 82:9546-9554.
Kerlan C, Moury B. 2008. Potato virus Y. Pages 287-296 in: Encyclopedia of Virology, Third Edition. Mahy BWJ, van Regenmortel MHV, eds. Elsevier, Oxford, UK.
Kimura M. 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press.
Knudsen B, Miyamoto MM. 2001. A likelihood ratio test for evolutionary rate shifts and functional divergence among proteins. Proc Natl Acad Sci U S A. 98:14512-14517.
Kryazhimskiy S, Plotkin JB. 2008. The population genetics of dN/dS. PLoS Genet. 4:e1000304.
Liang X.-Z, Lee BTK, Wong S-M. 2002. Covariation in the capsid protein of Hibiscus chlorotic ringspot virus induced by serial passaging in a host that restricts movement leads to avirulence in its systemic host. J Virol. 76:12320-12324.
López-Moya JJ, Pirone TP. 1998. Charge changes near the N terminus of the coat protein of two potyviruses affect virus movement. J Gen Virol. 79:161-165.
López-Moya JJ, Wang RY, Pirone TP. 1999. Context of the coat protein DAG motif affects potyvirus transmissibility by aphids. J Gen Virol. 80:3281–3288.
Martin DP, Williamson C, Posada D. 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21:260-262.
Meshi T, Motoyoshi F, Adachi A, Watanabe Y, Takamatsu N, Okada Y. 1988. Two concomitant base substitutions in the putative replicase genes of tobacco mosaic virus confer the ability to overcome the effects of a tomato resistance gene, Tm-1. EMBO J. 7:1575-1581.
Michaels SD, Amasino RM. 1998. A robust method for detecting single-nucleotide changes as polymorphic markers by PCR. Plant J. 14:381-385.
Moury B. 2004. Differential selection of genes of Cucumber mosaic virus subgroups. Mol Biol Evol. 21:1602–1611.
Moury B. 2010. A new lineage sheds light on the evolutionary history of Potato virus Y. Mol Plant Pathol 11:161-168.
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.
Moury B, Morel C, Johansen E, Guilbaud L, Souche S, Ayme V, Caranta C, Palloix A, Jacquemond M. 2004. Mutations in Potato virus Y genome-linked protein determine virulence toward recessive resistances in Capsicum annuum and Lycopersicon hirsutum. Mol Plant-Microbe Interact. 17:322-329.
Moury B, Fabre F, Senoussi R. 2007. Estimation of the number of virus particles transmitted by an insect vector. Proc Natl Acad Sci U S A. 104:17891-17896.
Neff MM, Neff JD, Chory J, Pepper AE. 1998. dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: Experimental applications in Arabidopsis thaliana genetics. Plant J. 14:387-392.
Ng J, Perry KL. 1999. Stability of the aphid transmission phenotype in cucumber mosaic virus. Plant Pathol. 48:388-394.
Ng JCK, Liu SJ, Perry KL. 2000. Cucumber mosaic virus mutants with altered physical properties and defective in aphid vector transmission. Virology 276:395-403.

Nielsen R. 2005. Molecular signatures of natural selection. Annu Rev Genet. 39:197-218

Nozawa M, Suzuki Y, Nei M. 2009. Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. Proc Natl Acad Sci U S A. 106:6700-6705.

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. Vir Res. 131:199–212.

Peng Y-H, Kadoury D, Gal-On A, Huet H, Wang Y, Raccah B. 1998. Mutations in the HC-Pro gene of zucchini yellow mosaic potyvirus: Effects on aphid transmission and binding to purified virions. J Gen Virol. 79:897-904.

Perry KL, Zhang L, Palukaitis P. 1998. Amino acid changes in the coat protein of cucumber mosaic virus differentially affect transmission by the aphids Myzus persicae and Aphis gossypii. Virology 242:204–210.

Pinel-Galzi A, Rakotomalala M, Sangu E, et al. (14 co-authors). 2007. Theme and variations in the evolutionary pathways to virulence of an RNA plant virus species. PLoS Path. 3:1761-1770.

Pond SLK, Frost SDW. 2005a. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21:2531–2533.

Pond SLK, Frost SDW. 2005b. Not so different after all: A comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 22:1208-1222.

Pond SLK, Frost SDW, Grossman Z, Gravenor, MB, Richman DD, Brown AJL. 2006. Adaptation to different human populations by HIV-1 revealed by codon-based analyses. PLoS Comp Biol. 2:530-538.

Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-818.

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.

Schirmer A, Link D, Cognat V, Moury B, Beuve M, Meunier A, Bragard C, Gilmer D, Lemaire O. 2005. Phylogenetic analysis of isolates of beet necrotic yellow vein virus collected worldwide. J Gen Virol. 86:2897-2911.

Schubert J, Fomitcheva V, Sztangret-Wisniewska J. 2007. Differentiation of Potato virus Y strains using improved sets of diagnostic PCR-primers. J Virol Meth. 140:66-74.

Seo JK, Kang SH, Seo BY, Jung, JK, Kim KH. 2010. Mutational analysis of interaction between coat protein and helper component-proteinase of Soybean mosaic virus involved in aphid transmission. Mol Plant Pathol. 11:265-276.

Sin SH, McNulty BC, Kennedy GG, Moyer JW. 2005. Viral genetic determinants for thrips transmission of Tomato spotted wilt virus. Proc Natl Acad Sci U S A. 102:5168-5173.

Shriner D, Nickle DC, Jensen MA, Mullins JI. 2003. Potential impact of recombination on sitewise approaches for detecting positive natural selection. Genet Res. 81:115–121.[

Suzuki Y, Nei M. 2004. False-positive selection identified by ML-based methods: Examples from the Sig1 gene of the diatom Thalassiosira weissflogii and the tax gene of a human T-cell lymphotropic virus. Mol Biol Evol. 21:914–921.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 24:1596-1599.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

Wyckoff GJ, Malcom CM, Vallender EJ, Lahn BT. 2005. A highly unexpected strong correlation between fixation probability of nonsynonymous mutations and mutation rate. Tr Genet. 21:381–385.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586-1591.

Yang Z, Nielsen R, Goldman N, Pedersen AMK. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449.

Yokoyama S, Tada T, Zhang H, Britt L. 2008. Elucidation of phenotypic adaptations: Molecular analyses of dimlight vision proteins in vertebrates. Proc Natl Acad Sci U S A. 105:13480–13485.

Supplementary Material

Accession numbers of PVY coat protein-coding region sequences used for phylogenetic analyses

AB025415, AB025416, AB042811, AB042812, AB042813, AB295476, AB295477, AB295479, AB331516, AB331519, AB331540, AB331544, AB331546, AB331548, AB331549, AB331550, AF012026, AF012027, AF012028, AF012029, AF255659, AF255660, AF321554, AF325927, AF463399, AF525081, AJ005639, AJ133454, AJ223592, AJ223593, AJ223594, AJ223595, AJ303093, AJ303094, AJ303095, AJ303096, AJ303097, AJ390289, AJ390290, AJ390292, AJ390293, AJ390294, AJ390295, AJ390297, AJ390298, AJ390299, AJ390300, AJ390301, AJ390302, AJ390303, AJ390304, AJ390305, AJ390306, AJ390307, AJ390309, AJ439544, AJ439545, AJ488834, AJ535662, AJ584851, AJ585195, AJ585196, AJ585197, AJ585198, AJ585342, AJ890342, AJ890343, AJ890344, AM236797, AM236814, AM411502, AY166866, AY166867, AY512655, AY601681, AY742726, AY742733, AY745491, AY745492, AY792597, AY841262, AY841265, AY884982, AY884983, AY884984, AY884985, D12539, DQ008213, DQ157178, EF026075, EF027861, EF027862, EF027863, EF027866, EF027897, EF027898, EU073856, EU252529, M81435, M95491, NC_001616, S74810, U09509, U10378, U25672, X14136, X54058, X54611, X68222, X68223, X68224, X68225, X68226, X79305, X92078, X97895, Z70237, Z70238, Z70239

Supplementary Table S1. Primers used for dCAPS analysis.

Primer name	Polarity	Primer sequence (5' to 3')	Position [a]
dCAPS-N25I	viral	GCAAAACAAGAGCAAGGTAGCATTCAACGAA [b]	8614-8644
dCAPS-E68K	viral	CCAAGAGTAAAGGTGCAACTGTACTAAATTTA [c]	8741-8772
Reverse	complementary	TCAAACTGTGATTGAGTTGC	8824-8843

[a] referring to isolate N605 (accession number X97895).
[b] the mismatch creating the *Eco*RI restriction site specific for the N605-N25I mutant is underlined.
[c] the mismatch creating the *Mse*I restriction site specific for the N605-E68K mutant is underlined.

Table 1: Codon positions of the coat protein coding region of two PVY groups significantly (0.95 *p*-value threshold) affected by positive selection according to the PAML software (Yang 2007) and three methods implemented in the HyPhy software (Pond and Frost 2005a). No codon position under positive selection was detected for clade C of PVY.

PVY group	PAML		HyPhy					
	Model M2a[a]		SLAC		FEL		REL	
	codon	*P*-value[b]	codon	*P*-value	codon	*P*-value	codon	*P*-value
Clade N + N×O recombinant clade (nucleotides 1 to 600)	25	0.998					25	0.987
							26	0.969
	68	0.996					68	0.998
							139	0.975
Clade O							9	0.988
	10	0.997			10	0.994	10	1.000
	11	0.955					11	1.000
							15	1.000
							16	1.000
							24	0.995
	26	0.996	26	0.952	26	0.979	26	1.000
							30	0.994
							138	0.994

[a] For each PVY group, model M2a performed better than models M0, M1a or M3 in likelihood ratio tests (LRTs). Model M8 also performed better than model M7 in a LRT and provided results similar to M2a.

[b] Posterior probability that individual codon positions belong to the positively-selected category using the Bayes Empirical Bayes method implemented in PAML (Yang 2007).

Table 2: Codon and nucleotide positions in the coat protein coding region of PVY with significant (*p*-value < 5%) evolutionary rate shifts[a] between PVY clades[b].

Fast group	Slow group	Nucleotide position	Codon position	*P*-value
C	O	74	25	0.0125
N	C	77	26	0.0212
N	O	74	25	0.0036
		75	25	0.0406
O	C	28	10	0.0194
		77	26	0.0022
Recombinant	C	202	68	0.0029
Recombinant	O	74	25	0.0063
		202	68	0.0273

[a] Likelihood ratio tests were performed according to Knudsen and Miyamoto (2001) on codon positions detected as candidates for positive selection using both PAML and HyPhy (Table 1).
[b] Clade N and the recombinant clade were not compared because they do not constitute two separate clades for this region of the coat protein coding region, a prerequisite for Knudsen and Miyamoto's (2001) method.

Table 3: Frequency of infection of tobacco plants (cultivar Xanthi) after aphid (*Myzus persicae* or *Aphis gossypii*) inoculation by the N605 PVY isolate or two coat protein mutants. Sources used for virus acquisition were Xanthi plants infected by each PVY variant.

	Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 4
Virus	*Myzus persicae*	*Myzus persicae*	*Myzus persicae*	*Myzus persicae*	*Aphis gossypii*
N605	11/105[a]	19/100	21/90	28/105	17/100
N605-CP25	14/105	28/100	19/90	25/105	nt[b]
N605-CP68	25/105** [c]	33/100*	44/90***	57/105***	32/100**

[a] Number of PVY infected plants/Total number of inoculated plants.
[b] nt: not tested.
[c] *, **, ***: significantly different from N605 at the 5%, 1% or 0.1% type-I error thresholds, respectively (Fisher exact test).

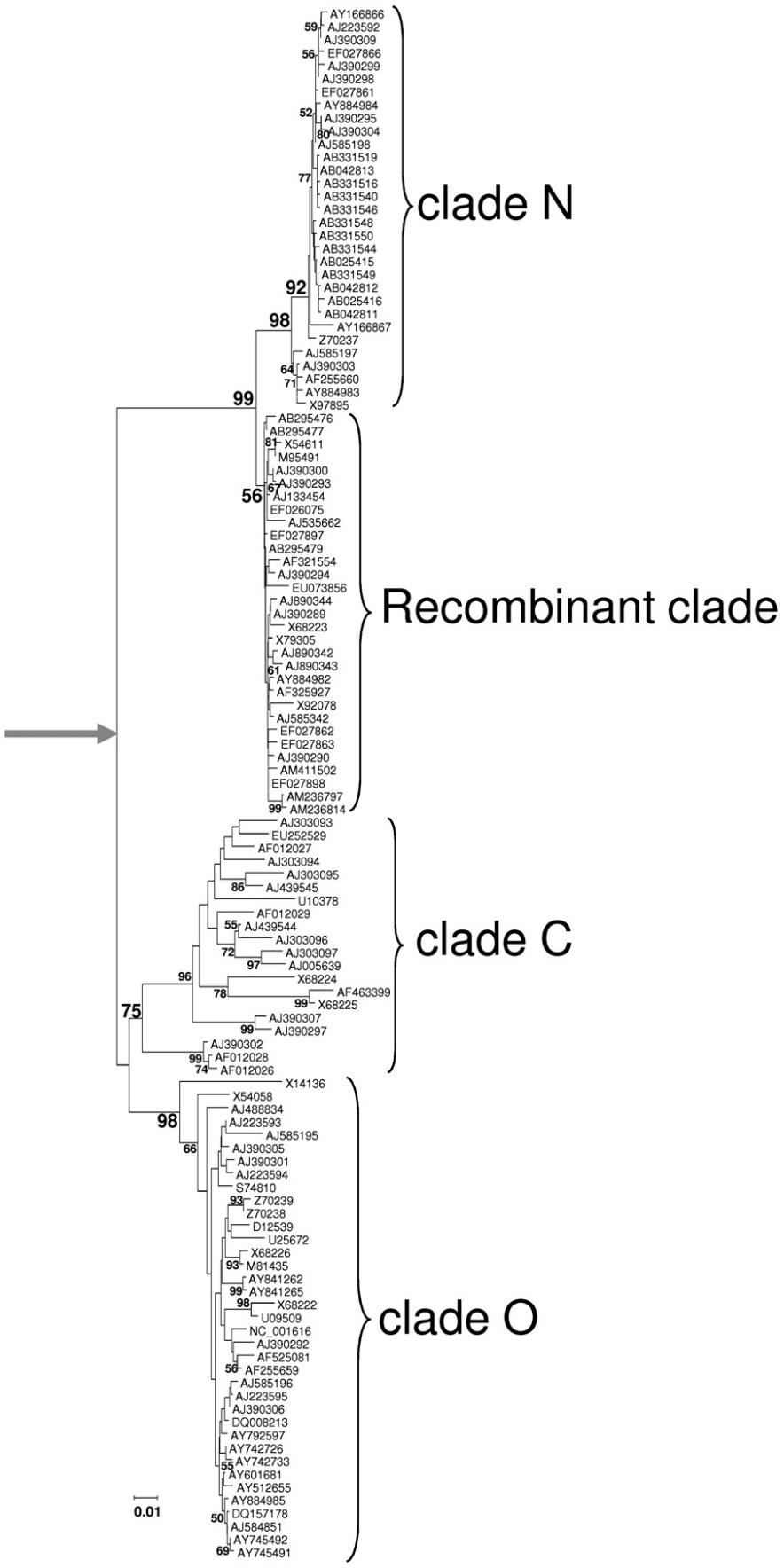

Fig. 1: Neighbor joining phylogenetic tree of the CP coding region of PVY isolates available in GenBank (December 2008). Bootstrap percentages above 50% are shown. The root of the tree is indicated by a gray arrow (Moury 2010). The scale bar indicates branch lengths in substitutions per nucleotide.

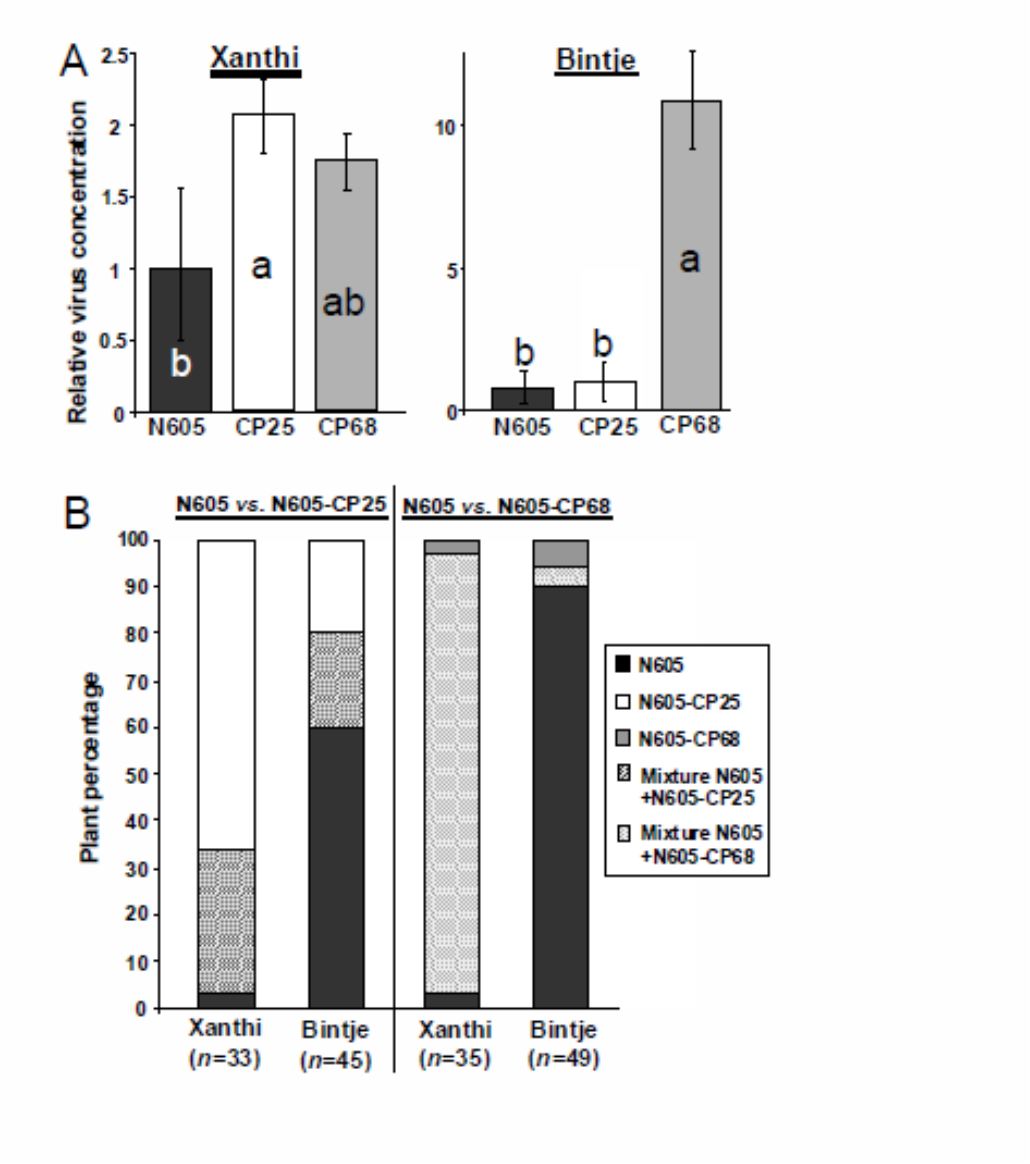

Fig. 2: Accumulation and competitiveness of three PVY variants in apical, non-inoculated leaves of the tobacco cultivar Xanthi and of the potato cultivar Bintje, 30 days post inoculation.
A: Virus accumulation in absence of competition between variants. The virus concentration relative to N605 (arbitrary value of 1.0) was assessed by DAS-ELISA and corrected for differential serological reactivity (see text). For each plant genotype, virus variants that have a letter in common do not show a significant difference in accumulation (Kruskal-Wallis test).
B: Competitiveness of the coat protein (CP) mutants against N605. *n* plants were inoculated manually with an inoculum composed of 50% of N605 and 50% of one of the CP mutants. The percentage of plants infected by N605, the CP mutant or both was assessed by dCAPS analysis.