Molecular evolution and phylogeography of *Potato virus Y* based on the CP gene

José M. Cuevas[1*], Agnès Delaunay[2**], Matevz Rupar[3], Emmanuel Jacquot[2**], Santiago F. Elena[1,4]

[1] Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-Universidad Politécnica de Valencia, València, Spain.

[2] INRA-Agrocampus Ouest-Université Rennes 1, UMR1099 BiO3P (Biology of Organisms and Populations Applied to Plant Protection), Le Rheu, France.

[3] NIB, Department of Biotechnology and Systems Biology, Vecna pot 111, SI-1000 Ljubljana, Slovenia.

[4] The Santa Fe Institute, Santa Fe NM, United States of America.

*Corresponding author: José M. Cuevas, Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-UPV, Campus UPV CPI 8E, Ingeniero Fausto Elio s/n, 46022 València, Spain. Phone: 34 963 877 895, FAX: 34 963 877 859, e-mail: cuevast@uv.es.

**: present address: INRA-CIRAD-Supagro, UMR385 BGPI (Biology and Genetic of Plant-Pathogen Interactions), Montpellier, France

Running title: PVY CP molecular evolution

2590 words in the main text, 146 words in the summary, 3 main tables, 6 supplementary tables and 1 supplementary figure.

KEYWORDS: RNA virus, genetic variability, positive selection, covariation, virus evolution.

Footnote page: The authors do not have a commercial or any other association that might pose a conflict of interest.

Summary

Potato virus Y (PVY) is an important plant pathogen with a wide host range that includes, among others, potato, tobacco, tomato, and pepper. The coat protein (CP) of PVY has been commonly used in phylogenetic studies for strain classification. In this study, we used a pool of 292 CP sequences from isolates collected worldwide. After detecting and removing recombinant sequences, we applied Bayesian techniques to study the influence of geography and host species in CP population structure and dynamics. Finally, we performed selection and covariation analyses to identify specific amino acids involved in adaptation. Our results show that PVY CP diversification is significantly accounted for by both geographic and host-driven adaptations. Amino acid positions detected as positively selected concentrate in the N-terminal region of the protein. Some of these selected positions may discriminate among strains, and to a much lesser extent, between potato and non-potato isolates.

Potato potyvirus Y (PVY) is responsible for serious diseases in potato, tobacco, pepper, and tomato crops. PVY was originally classified into strain groups (e.g., PVY^N, PVY^O and PVY^C) according to biological properties, serological characteristics and/or genome sequences (Moury *et al.*, 2002; Singh *et al.*, 2008). Recombination is highly pervasive in PVY and additional genomic organizations have been recently described (Lorenzen *et al.*, 2008; Schubert *et al.*, 2007).

Molecular evolution studies are useful tools to shed light on the molecular bases of virus geographical spread and adaptation to new hosts and for designing better epidemics control strategies (Elena *et al.*, 2011; Jones, 2009). We recently studied the phylogeography and molecular evolution of PVY whole-genomes (Cuevas *et al.*, 2012), showing that host and geographic origin influenced PVY diversification, and detecting positively selected sites. Here we revisit these topics but focusing on the CP. Novelties of this study are: i) a much larger data set is available for the CP, which is expected to allow a more robust characterization of phylogenetic and selection patterns, ii) the CP plays an important role in host adaptation for many plant viruses, and iii) the CP is the most diverse and well-studied gene in PVY and other potyviruses (Moury & Simon, 2011; Ogawa *et al.*, 2008; Rohozkova & Navratil, 2011; Visser & Bellstedt, 2009).

A detailed description of the methods employed in this study can be found elsewhere (e.g., Cuevas *et al.*, 2012). For this study, we retrieved 198 PVY CP sequences from GeneBank, plus 94 additional sequences from worldwide isolates (PVYwide Organization, http://www.inra.fr/pvy_organization) (Table S1). This dataset was aligned with MUSCLE (Edgar, 2004) as implemented in MEGA 5 (Tamura *et al.*, 2011). We run recombination analyses to remove its effect from subsequent analyses. Bayesian Markov chain Monte Carlo (MCMC) coalescent analyses were performed with non-recombinant isolates to study the effect of local adaptation and host species in the observed diversity. Finally, we performed selection analyses to identify regions from the CP cistron that may be more likely involved in PVY adaptation dynamics.

Seventy-five out of the 292 isolates (Table S1) showed a breakpoint indicating ancestral recombination between PVY^N and PVY^O strains at position 9170 (considering the full genome) in the CP (Schubert *et al.*, 2007) and worldwide distributed. Five other isolates showed uncommon breakpoints detected by at least three of the methods implemented in RDP3 (Martin *et al.*, 2010). N Nysa isolate showed a newly described breakpoint at position 8896 (Cuevas *et al.*, 2012). IAC and v951204-N isolates showed a breakpoint at position 8735 (being Mont and SASA-110 the major and minor parents, respectively), almost coincident with other previously described breakpoints (Moury *et al.*, 2002). Finally, S-RB96 and NN-UK-N isolates showed a new recombination point at position 8947 (SASA-110 and Mont are the major and minor parents, respectively). All recombinants were excluded, reducing the dataset to 212 isolates.

Phylogenetic analyses were performed using the GTR + Γ_4 + I substitution model in the Bayesian MCMC framework, as implemented in BEAST 1.6 (Drummond & Rambaut, 2007). Substitution rates were estimated using the relaxed uncorrelated exponential clock model. The three typical PVY strain groups (PVY^C, PVY^O and PVY^N) could be observed (Figure S1), although the differentiation between PVY^C and PVY^O strains was poorly supported. Chile3 occupies a basal position in the tree, outside any of the strain groups, supporting its ancestry (Moury, 2010). Within the PVY^C clade, 17 out of 22 isolates were collected from five different non-potato hosts. However, host species did not account for clustering within this clade, since most of the isolates from a given host were dispersed along the clade or closely grouped with isolates from other hosts. Only isolates PVY-MN and NC57 (from tobacco) formed a differentiated cluster, as previously observed (Kehoe & Jones, 2011; Mascia *et al.*, 2010). PVY^C clade has been subdivided into PVY^{C1} and PVY^{C2} subgroups depending on their ability to infect pepper (Blanco-Urgoiti *et al.*, 1998). In our phylogenetic tree, only isolates PVY-C-CM and Adgen-C were of pathotype PVY^{C2}, forming a differentiated cluster. Isolate CAA82 collected from pepper, grouped outside the PVY^{C1} subgroup. More isolates from subgroup PVY^{C2} are thus necessary to check the relative distance of isolate

CAA82 to those from non-pepper subgroup PVY^{C2}. Most isolates in our data set belong to PVY^{O}. The globally low branch supports suggests a very genetically homogeneous group, compatible with a recent origin with minimal selection (Pagán *et al.*, 2006; Roossinck *et al.*, 1999). In fact, well-supported clusters within the PVY^{O} clade included isolates with common geographic origins. Finally, a similar trend was observed in the PVY^{N} clade, although internal branches close to the basis of the tree were usually well supported, thus differentiating several monophyletic clusters. Our study supports the classification proposed by Ogawa *et al*. (2008) into two PVY^{N} main groups (i.e., N-Europe and N-North America). Some well-supported clusters were observed into each PVY^{N} group, although this differentiation was not strictly associated with geographic origin.

A visual inspection of the maximum clade credibility (MCC) phylogeny did not show a clear structure in terms of geographic origin at the continent level (Figure S1 and Table S1). For commercial and geographical reasons, North African and Middle East isolates were included into the European group. For the same reason, the only isolate from New Zealand was not included into any continental group. We used BATS 1.0b2 (Parker *et al.*, 2008) to calculate three statistics (*AI:* association index, *PS:* parsimony score and *MC:* maximum monophyletic clade size) describing the correlation between the geographic and the phylogenetic relationships. Significant signatures for geographic structure in the diversity of CP cistron were observed when grouped by geographic origins (Table 1), as shown by the significant *AI* and *PS* values. Asian, European, South African, and North American groups showed differentiated subpopulations (significant *MC* values). South American group did not show a significant association, which is accounted for by the small sample size, and no inference was possible for the single New Zealand isolate.

Host-driven adaptation could also be tested using host as grouping variable, and a significant signature was also observed (Table 1). In this case, the differentiation was due to three subpopulations of isolates derived from potato, tobacco and pepper. For tomato and black nightshade no significant association was detected, whereas no inference was possible for single isolates from ají and tamarillo. Since most of the samples in our data set are potato isolates, the significance of *AI* and *PS* values could be a consequence of the global distribution of the same state across most of the branches in the tree (Parker *et al.*, 2008). However, host structure explained quite well the phylogeny, since clade PVY^{C} predominantly included non-potato isolates (17 out of 22), whereas the remaining main clades only included 14 non-potato isolates (out of 189). Twelve out of the 14 non-potato isolates falling outside the PVY^{C} clade were collected from tobacco. In this sense, tobacco infection could accidentally take place from potato crops early in the year, thus leading to misidentification of some tobacco isolates (M. Chrzanowska pers. comm.). Besides, it is not surprising either that tomato isolate GR_PVY12 fell outside clade PVY^{C}, since tomato can be infected with most PVY potato isolates (Singh *et al.*, 2008), and thus a recent introduction from potatoes cannot be excluded. Finally, the inclusion of black nightshade isolate SYR-Sn into PVY^{O} clade is surprising, although the biological properties of this isolate are not yet available.

Selective pressures at a codon level were estimated using FEL, IFEL and MEME methods (www.datamonkey.org). Intramolecular covariation analyses were carried out using CAPS 1 (Fares & Travers, 2006), as previously described (Cuevas *et al.*, 2012). Table 2 shows the distribution of codon positions under purifying, neutral and positive selection, and covarying positions. As previously shown, most of the codons evolve neutrally, whereas purifying selection is the main force driving the evolution of CP (Cuevas *et al.*, 2012). Negatively selected positions are scattered along the ORFs, suggesting that no domain is particularly constrained. FEL and IFEL predicted codon one as positively selected, whereas MEME detected three additional codons (68, 193 and 216) to be under episodic diversifying selection (Table 2). Finally, a covariation group of nine codons was also detected, all located at the first half of the CP. Selected codon one was involved into this covariation group.

Previous phylogenetic studies showed that non-potato isolates mainly fell into clade PVY^{C} (Ogawa *et al.*, 2008; Schubert *et al.*, 2007), highlighting the importance of host-driven adaptation. Our study, which included a significantly larger number of non-potato isolates, clearly showed that, in spite of the global consideration of non-potato isolates as belonging to the clade PVY^C, several other non-potato isolates were dispersed in the phylogeny. In fact, the analysis of amino acid composition for positively selected and covarying positions showed no clear differences between potato and non-potato isolates (Tables S2 and S3). Globally, both groups, except for positions 24, 138 and 193, shared the same predominant amino acid at a given position. Whereas similar amino acid composition between both groups was found for positions 24 and 193, the main difference was found at position 138, since the predominant amino acid for non-potato isolates was absent in potato isolates (Table S3). Besides, with the exception of position 138, specific residues of potato and non-potato isolates were always present at low frequencies. We also obtained the amino acid composition of positively selected and covarying codons, but grouping in this case for the PVY^C, PVY^O and PVY^N strains, which allowed us to check if selective forces were strain-specific (Tables S4 and S5). Globally, the same predominant amino acid at a given position was usually shared by the three strains. For those cases showing differences in the predominant amino acid, these predominant residues for a given strain were also usually present at low frequencies in at least one of the alternative strains. We observed positions 24 and 193 wherein the predominant amino acid for PVY^O strain was different from that of PVY^C and PVY^N strains. Besides, the predominant amino acid from PVY^N strain was different from that observed at PVY^C and PVY^O strains for positions 1, 11, 17, 26, 29, and 31. Finally, positions 99 and 138 showed different predominant residues for the three strains. Interestingly, the predominant residue for the PVY^C strain at these two positions was absent in the other two strains, although the predominant amino acids from PVY^O and PVY^N strains were also present at low frequencies. Consequently, the analysis of amino acid composition at selected and covarying positions showed more partially discriminant residues among strains than among potato and non-potato isolates, which indicates that selective forces are mainly acting independently of the potato/non-potato distinction. In this sense, as mentioned before, PVY does not have a narrow host range, which would account for the lack of association between selected positions and host usage.

Selection analyses at a branch level were performed using SWAPSC (Fares, 2004) to check the potential association between selective events and the phylogeny. Thirty-four branches showed evidence of positive selection (18 internal and 16 terminal branches; Figure S1), and this selective signature was detected in 13 regions, often overlapping (Table 3). Most of them fell into the N-terminal region, congruently with the above selection and covariation analyses (Tables 2 and 3). Respect to the distribution of the selected branches in the phylogeny, we could differentiate between internal and terminal branches (Figure S1). The frequency of selected internal branches was different among clades (20%, 3.7% and 15.8% for PVY^C, PVY^O and PVY^N clades, respectively; Fisher's exact test, $P = 0.003$), but not for terminal branches (with frequencies of 9.1%, 6.1%, and 10.5% for PVY^C, PVY^O and PVY^N clades, respectively; Fisher's exact test, $P = 0.568$). These results suggest that selective forces are stronger into the PVY^C and PVY^N clades and milder into PVY^O. It is worth mentioning that one selected internal branch lead to PVY^C clade (named as b2 in Table 3 and Figure S1), except for the tamarillo isolate falling outside the selected cluster. We obtained the amino acid composition of the region involved in this branch specific selection event (codons 187-194) for PVY^C, PVY^O and PVY^N clades (Table S6). This region included selected site 193, which have been discussed above. Besides, the predominant amino acid for PVY^N clade was different from that observed at PVY^C and PVY^O clades at position 187. Finally, position 194 clearly discriminated between PVY^O and PVY^N clades, but the two fixed residues present in these strains were also observed in the PVY^C strain. In conclusion, branch selection analyses showed evidence of the differential effect of selective events among strains, but did not provide particular positions accounting for these differences at a strain level.

The role of CP protein in the pathology of potyviruses have been previously confirmed (Andrejeva *et al.*, 1999; Hu *et al.*, 2011; Ullah & Grumet, 2002) and symptom determinants may be different even between strains of PVY in a particular host (Bukovinszki *et al.*, 2007). The N-terminal part of CP protein is a clear example of multifunctionality. It is exposed on the virion surface (potential function in binding ligands), besides being involved in vector transmission (Peng *et al.*, 1998) and systemic plant colonization (Andersen & Johansen, 1998; López-Moya & Pirone, 1998), becoming a potential target of selection at both vector and plant levels. In addition, CP protein from PVY interacts with different chloroplast proteins (Feki *et al.*, 2005). Consequently, it is not easy to discern if a given amino acid position is involved into one or more functions.

Regarding biological functions of the CP protein, several commonalities were found when comparing our results with those described by Moury and Simon (2011). All positions showing positive selection in this previous study are within the N-terminal region of the CP cistron. In particular, positions 11, 24, 26, 68, and 138, were also detected to be under positive selection or covariation in our study. Position 11 is close to the DAG conserved motif involved in aphid transmission (Atreya *et al.*, 1991, 1995), and it has been shown that mutations in a neighbor residue can reduce substantially transmissibility (Atreya *et al.*, 1995). Furthermore, position 25 was shown to affect virus accumulation in host plants (Moury & Simon, 2011), and covarying positions detected in the vicinity could have some influence in this respect. Regarding position 68, it is worth mentioning that a mutation in this codon promoted differences in viral accumulation and transmissibility by aphids (Moury & Simon, 2011). Finally, the region spanning amino acid positions 133 to 148 of the CP from *Soybean mosaic virus* (positions 136-151 of PVY CP), is involved in binding to the HC-Pro (Seo *et al.*, 2010), and then a potential influence for the included covarying position 138 could be postulated.

Acknowledgements

This research was supported by the Spanish Ministry of Science and Innovation grant BFU2009-06993 to SFE. JMC was supported by the JAE-Doc program from CSIC. We thank Mario A. Fares for fruitful bioinformatics guidance. PVY isolates used in this work were available thanks to PVYwide Organization members (Carole Balmelli, Peter Dedic, Stewart Gray, Christophe Lacomme,

Kestin Lindner, Benoit Moury, Maja Ravnikar, Laura Tomassoli, René Van der Vlugt, and Christina Varaveri).

References

Andersen, K. & Johansen, I. E. (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, J. P. (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.

Atreya, P. L., Atreya, C. D. & Pirone, T. P. (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, P. L., López-Moya, J. J., Chu, M., Atreya, C. D. & Pirone, T. P. (1995). Mutational analysis of the coat protein N-terminal amino acids involved in potyvirus transmission by aphids. *J Gen Virol* **76**, 265-270.

Blanco-Urgoiti, B., Sanchez, F., Perez de San, R. C., Dopazo, J. & Ponz, F. (1998). Potato virus Y group C isolates are a homogeneous pathotype but two different genetic strains. *J Gen Virol* **79**, 2037-2042.

Bukovinszki, A., Gotz, R., Johansen, E., Maiss, E. & Balazs, E. (2007). The role of the coat protein region in symptom formation on *Physalis floridana* varies between PVY strains. *Virus Res* **127**, 122-125.

Cuevas, J. M., Delaunay, A., Visser, J. C., Bellstedt, D. U., Jacquot, E. & Elena, S. F. (2012). Phylogeography and molecular evolution of Potato virus Y. *PLoS ONE* **7**, e37853.

Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214.

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

Elena, S. F., Bedhomme, S., Carrasco, P., Cuevas, J. M., de la Iglesia, F., Lafforgue, G., Lalic, J., Pròsper, A., Tromas, N. & Zwart, M. P. (2011). The evolutionary genetics of emerging plant RNA viruses. *Mol Plant Microbe Interact* **24**, 287-293.

Fares, M. A. (2004). SWAPSC: sliding window analysis procedure to detect selective constraints. *Bioinformatics* **20**, 2867-2868.

Fares, M. A. & Travers, S. A. (2006). A novel method for detecting intramolecular coevolution: adding a further dimension to selective constraints analyses. *Genetics* **173**, 9-23.

Feki, S., Loukili, M. J., Triki-Marrakchi, R., Karimova, G., Old, I., Ounouna, H., Nato, A., Nato, F., Guesdon, J.-L., Lafaye, P. & Ben Ammar Elgaaied, A. (2005). Interaction between tobacco Ribulose-1,5-biphosphate Carboxylase/Oxygenase large subunit (RubisCO-LSU) and the PVY Coat Protein (PVY-CP). *Eur J Plant Pathol* **112**, 221-234.

Hu, X., Nie, X., He, C. & Xiong, X. (2011). Differential pathogenicity of two different recombinant PVY(NTN) isolates in *Physalis floridana* is likely determined by the coat protein gene. *Virol J* **8**, 207.

Jones, R. A. (2009). Plant virus emergence and evolution: origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. *Virus Res* **141**, 113-130.

Kehoe, M. A. & Jones, R. A. (2011). A proposal to help resolve the disagreement between naming of Potato virus Y strain groups defined by resistance phenotypes and those defined by sequencing. *Arch Virol* **156**, 2273-2278.

López-Moya, J. J. & Pirone, T. P. (1998). Charge changes near the N terminus of the coat protein of two potyviruses affect virus movement. *J Gen Virol* **79**, 161-165.

Lorenzen, J., Nolte, P., Martin, D., Pasche, J. S. & Gudmestad, N. C. (2008). NE-11 represents a new strain variant class of Potato virus Y. *Arch Virol* **153**, 517-525.

Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D. & Lefeuvre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**, 2462-2463.

Mascia, T., Finetti-Sialer, M. M., Cillo, F. & Gallitelli, D. (2010). Biological and molecular characterization of a recombinant isolate of Potato virus Y associated with a tomato necrotic disease occurring in Italy. *J Plant Pathol* **92**, 131-138.

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. & Simon, V. (2011). dN/dS-based methods detect positive selection linked to trade-offs between different fitness traits in the coat protein of Potato virus Y. *Mol Biol Evol* **28**, 2707-2717.

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.

Pagán, I., Carmen Córdoba-Sellés, M., Martínez-Priego, L., Fraile, A., Malpica, J. M., Jordá, C. & García-Arenal, F. (2006). Genetic Structure of the Population of Pepino mosaic virus Infecting Tomato Crops in Spain. *Phytopathology* **96**, 274-279.

Parker, J., Rambaut, A. & Pybus, O. G. (2008). Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect Genet Evol* **8**, 239-246.

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.

Rohozkova, J. & Navratil, M. (2011). P1 peptidase--a mysterious protein of family Potyviridae. *J Biosci* **36**, 189-200.

Roossinck, M. J., Zhang, L. & Hellwald, K. H. (1999). Rearrangements in the 5' nontranslated region and phylogenetic analyses of Cucumber mosaic virus RNA 3 indicate radial evolution of three subgroups. *J Virol* **73**, 6752-6758.

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.

Seo, J. K., Kang, S. H., Seo, B. Y., Jung, J. K. & Kim, K. H. (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.

Singh, R. P., Valkonen, J. P., Gray, S. M., Boonham, N., Jones, R. A., Kerlan, C. & Schubert, J. (2008). Discussion paper: The naming of Potato virus Y strains infecting potato. *Arch Virol* **153**, 1-13.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-2739.

Ullah, Z. & Grumet, R. (2002). Localization of Zucchini yellow mosaic virus to the veinal regions and role of viral coat protein in veinal chlorosis conditioned by the *zym* potyvirus resistance locus in cucumber. *Physiol Mol Plant Pathol* **60**, 79-89.

Visser, J. C. & Bellstedt, D. U. (2009). An assessment of molecular variability and recombination patterns in South African isolates of Potato virus Y. *Arch Virol* **154**, 1891-1900.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591.

Table 1. Analysis of the geographic and host effect on the population structure of PVY isolates.

Analyses	# Isolates	Association	Test value	*P*
Geographic		*PS*	106.985	<0.001
		AI	17.979	<0.001
Asia	30	*MC*	1.911	0.0099
Europe	88	*MC*	3.623	0.0099
South Africa	47	*MC*	2.451	0.0099
North America	43	*MC*	2.272	0.0099
South America	3	*MC*	1.004	1
New Zealand	1	*MC*	NA[1]	
Host species		*PS*	31.480	<0.001
		AI	6.651	<0.001
Potato	180	*MC*	13.145	0.0199
Tobacco	14	*MC*	1.286	0.0400
Pepper	10	*MC*	1.136	0.0099
Tomato	4	*MC*	1.005	1
Black nightshade	2	*MC*	1.001	1
Ají	1	*MC*	NA[1]	
Tamarillo	1	*MC*	NA[1]	

[1]insufficient sample size ($n < 2$).

Table 2. Results of the codon selection and covariation analyses at the CP gene. For selection methods (FEL, IFEL and MEME), the number of codons detected to be under negative, neutral or positive selection are given. The last column indicates the location of positively selected sites besides those positions showing covariation (CAPS).

	Negative	**Neutral**	**Positive**	**Location**
FEL	113	153	1	1
IFEL	76	190	1	1
MEME	NA	NA	4	1, 68, 193, 216
CAPS	-	-	-	1, 11, 17, 24, 26, 29, 31, 99, 138

Table 3. Results of branch selection analysis. First column indicates all regions (codons) showing evidence of positive selection and second column shows the branches associated with the selection event for a given region. For terminal branches, the name of the corresponding isolate is shown. Internal branches are numbered as indicated in Figure S1 and marked in bold. Positively selected and covarying positions falling into the regions providing a positive selection signature are shown in the last two columns, respectively.

Region	Branch	FEL-IFEL-MEME	Covariation
7-11	SASA-110, **b3**		11
8-13	PN-82		11
23-28	**b12**		24, 26
23-29	**b14**		24, 26, 29
25-28	PB_707, US05_30, SYR-NB-16		26
25-29	**b15, b16, b17**		26, 29
26-29	**b5, b7**, **b11**		26, 29
29-33	CAA141, PB_707, PB_602, PB_752, SC143, SC61, US05_30, US05_7, NN71_111, SYR-NB-16, 605, **b4**, **b8**, **b9, b13**, **b18**		29, 31
62-65	German_45, US06_55, **b6**		
135-138	**b1**		138
187-193	Nicola	193	
187-194	**b2**, **b3**, **b10**	193	
214-217	German_45, **b6**	216	

Table S1. PVY isolates used in the present study.

Isolate	GenBank accession	Origin	Host	Collection date
156	AJ889867	Germany	*S.tuberosum*	
605	X97895	Switzerland	*S.tuberosum*	1976
12-94*	AJ889866	Poland	*S.tuberosum*	1994
156var	AJ889868	Germany	*S.tuberosum*	2004
261-4	AM113988	Germany	*S.tuberosum*	2004
34/01*	AJ890342	Poland	*S.tuberosum*	2001
423-3*	AY884982	USA	*S.tuberosum*	2002
53-29	AJ390298	Denmark	*S.tuberosum*	
53-49	AJ390299	Denmark	*S.tuberosum*	
Adgen-C	AJ890348	France	*S.tuberosum*	2005
Al-Baqa'*	EU073854	Jordan	*S.tuberosum*	
Al-Ghor*	EU073855	Jordan	*S.tuberosum*	
Al-Mafraq*	EU073857	Jordan	*S.tuberosum*	
aL-Ramtha	EU073859	Jordan	*S.tuberosum*	
Alt*	AY884985	USA	*S.tuberosum*	2002
Anqiu4	EF592517	China	*N.tabacum*	
AQ1	EF592513	China	*N.tabacum*	
Ca/H*	AJ535662	Hungary	*C.annuum*	
CAA141	JQ954317	France	*C.annuum*	1999
CAA15	JQ954318	France	*C.annuum*	2000
CAA82	JQ954315	Israel	*C.annuum*	1982
CAPA7	JQ954316	Tunisia	*C.annuum*	2006
CC24_5	GQ853667	South Africa	*S.tuberosum*	
CC55_8_146	GQ853652	South Africa	*S.tuberosum*	
CC62_20_156*	GQ853623	South Africa	*S.tuberosum*	
CC66_91_47	GQ853653	South Africa	*S.tuberosum*	
CC9_12_171*	GQ853621	South Africa	*S.tuberosum*	
CC9_30_175	GQ853650	South Africa	*S.tuberosum*	
CC9_47_177*	GQ853622	South Africa	*S.tuberosum*	
CC9_48_178	GQ853651	South Africa	*S.tuberosum*	
Chile3	FJ214726	Chile	*C.baccatum*	2005
DD019_141_138	GQ853661	South Africa	*S.tuberosum*	
DD020_92_30*	GQ853624	South Africa	*S.tuberosum*	
DD037F_31_186*	GQ853625	South Africa	*S.tuberosum*	
DD037F_35_188*	GQ853626	South Africa	*S.tuberosum*	
DD037F_9_154	GQ853655	South Africa	*S.tuberosum*	
DD051_14	GQ853660	South Africa	*S.tuberosum*	
DD051_7	GQ853659	South Africa	*S.tuberosum*	2007
DD103A_101_190	GQ853657	South Africa	*S.tuberosum*	
DD103A_184_191	GQ853603	South Africa	*S.tuberosum*	
DD103A_80_180	GQ853656	South Africa	*S.tuberosum*	
DD122A_25	GQ853658	South Africa	*S.tuberosum*	
DD122A_34	GQ853662	South Africa	*S.tuberosum*	

DD122A_36*	GQ853627	South Africa	*S.tuberosum*	
Ditta*	AJ890344	Austria	*S.tuberosum*	1998
Fanzhen6	EF592515	China	*S.tuberosum*	
Fanzhen8	EF592521	China	*S.tuberosum*	
Foggia	EU482153	Italy	*L.esculentum*	2007
FX24	EF592514	China	*S.tuberosum*	
German_14*	JQ954384	Germany	*S.tuberosum*	2003
German_16*	JQ954295	Germany	*S.tuberosum*	2003
German_20	JQ954387	Germany	*S.tuberosum*	2004
German_33	JQ954296	Germany	*S.tuberosum*	2004
German_34	JQ954297	Germany	*S.tuberosum*	2004
German_35	JQ954298	Germany	*S.tuberosum*	2004
German_37	JQ954299	Germany	*S.tuberosum*	2004
German_38	JQ954300	Germany	*S.tuberosum*	2004
German_39	JQ954301	Germany	*S.tuberosum*	2004
German_4*	JQ954302	Germany	*S.tuberosum*	2002
German_41	JQ954303	Germany	*S.tuberosum*	2004
German_42	JQ954304	Germany	*S.tuberosum*	2004
German_43	JQ954305	Germany	*S.tuberosum*	2004
German_45	JQ954306	Germany	*N.tabacum*	2006
German_47	JQ954314	Germany	*S.tuberosum*	2003
German_51	JQ954307	Germany	*S.tuberosum*	2004
German_52	JQ954308	Germany	*S.tuberosum*	2004
German_55	JQ954309	Germany	*S.tuberosum*	2004
German_56	JQ954310	Germany	*S.tuberosum*	2004
German_57	JQ954311	Germany	*S.tuberosum*	2004
German_58	JQ954312	Germany	*S.tuberosum*	2004
German_62*	JQ954342	Germany	*S.tuberosum*	2004
German_65	JQ954313	Germany	*S.tuberosum*	2004
GG517_128	GQ853635	South Africa	*S.tuberosum*	2005
GG517_170_168	GQ853636	South Africa	*S.tuberosum*	
GG517_93_160	GQ853593	South Africa	*S.tuberosum*	
Gpost*	JN936420	South Africa	*S.tuberosum*	2010
GR_PVY12	JQ954319	Greece	*L.esculentum*	1998
GR_PVY13	JQ954320	Greece	*L.esculentum*	1999
GR_PVY84*	JQ954321	Greece	*S.tuberosum*	2004
Gr99*	AJ890343	Poland	*N.tabacum*	1999
Hangzhou	AJ488834	China	*S.tuberosum*	
HN2	GQ200836	China	*S.tuberosum*	2007
Thole*	M95491	Hungary	*S.tuberosum*	1993
IAC**	AY840082	Brazil	*S.tuberosum*	
Irbid*	EU073856	Jordan	*S.tuberosum*	
Isol5	AJ890350	Germany	*S.tuberosum*	
IT_104	JQ954323	Italy	*S.tuberosum*	1997
IT_115*	JQ954324	Italy	*S.tuberosum*	1998
IT_117	JQ954325	Italy	*S.tuberosum*	1998

IT_101*	JQ954322	Italy	*S.tuberosum*	1998
IT_215*	JQ954326	Italy	*S.tuberosum*	1998
L26*	FJ204165	USA	*S.tuberosum*	2007
Laiwu1	EF592525	China	*S.tuberosum*	
Laiwu29	EF592527	China	*N.tabacum*	
Laiwu3	EF592516	China	*S.tuberosum*	
Laiwu9*	EF592526	China	*S.tuberosum*	
Linda*	AJ890345	Germany	*S.tuberosum*	2004
Linkou29	EF592524	China	*N.tabacum*	
LW	AJ890349	Poland	*S.tuberosum*	1970
LYE842	AJ43954	Canary Islands	*L.esculentum*	1984
ME173	FJ643479	USA	*S.tuberosum*	2006
Mengyin3	EF592518	China	*N.tabacum*	
MengyinA	EF592519	China	*N.tabacum*	
MengyinC	EF592520	China	*N.tabacum*	
Mont	AY884983	USA	*S.tuberosum*	2001
N_Nysa**	FJ666337	Poland	*S.tuberosum*	1974
N484_1	GQ853634	South Africa	*S.tuberosum*	
Naur*	EU073858	Jordan	*S.tuberosum*	
NC57	DQ309028	USA	*N.tabacum*	1973
NE-11*	DQ157180	USA	*S.tuberosum*	2003
N-Egypt	AF52229	Egypt	*S.tuberosum*	2001
New_Zealand	AM268435	New Zealand	*S.tuberosum*	2002
NIB-NTN*	AJ585342	Slovenia	*S.tuberosum*	
Nicola	AJ890346	Germany	*N.tabacum*	1999
N-Jg	AY166867	Canada	*S.tuberosum*	1991
NN300_155_19	GQ853597	South Africa	*S.tuberosum*	
NN300_155_22	GQ853598	South Africa	*S.tuberosum*	
NN300_41_123	GQ853595	South Africa	*S.tuberosum*	
NN300_60_23	GQ853663	South Africa	*S.tuberosum*	
NN300_76_118*	GQ853628	South Africa	*S.tuberosum*	
NN300_98_31	GQ853596	South Africa	*S.tuberosum*	
NN300_99_34	GQ853664	South Africa	*S.tuberosum*	
NN333B_28_149*	GQ853629	South Africa	*S.tuberosum*	
NN333B_87_152*	GQ853630	South Africa	*S.tuberosum*	
NN459_14	GQ853631	South Africa	*S.tuberosum*	
NN459_25	GQ853599	South Africa	*S.tuberosum*	
NN71_111	GQ853594	South Africa	*S.tuberosum*	2005
nnp*	AF237963	Italy	*C.annuum*	1992
NN-UK-N**	AJ390296	UK	*S.tuberosum*	
NN-UK-O	AJ390297	UK	*S.tuberosum*	
NTND6	AB331515	Japan	*S.tuberosum*	1997
NTNHO90	AB331517	Japan	*S.tuberosum*	1997
NTNHO92	AB331549	Japan	*S.tuberosum*	
NTNHO95	AB331550	Japan	*S.tuberosum*	
NTNK114	AB331540	Japan	*S.tuberosum*	

NTNNN99	AB331518	Japan	*S.tuberosum*	1997
NTNOK102	AB331546	Japan	*S.tuberosum*	
NTNOK105	AB331516	Japan	*S.tuberosum*	1997
NTNON92	AB331519	Japan	*S.tuberosum*	1997
O-Des	AJ390305	UK	*S.tuberosum*	
O-Gov	AJ390301	UK	*S.tuberosum*	
O-Tom	AJ390307	Portugal	*S.tuberosum*	
P21-82	AJ303097	Spain	*C.annuum*	
P21-82b	AJ005639	Spain	*C.annuum*	
PB_602	JQ954329	The Netherlands	*S.tuberosum*	1978
PB_702	JQ954327	The Netherlands	*S.tuberosum*	1957
PB_707	JQ954328	The Netherlands	*S.tuberosum*	1958
PB_752	JQ954330	The Netherlands	*S.tuberosum*	1995
PB312*	EF026075	USA	*S.tuberosum*	2003
PMB21	AJ390306	UK	*S.tuberosum*	
PN10A	DQ008213	USA	*S.tuberosum*	2004
PN-82	AJ303096	Spain	*C.annuum*	
PO7	U09509	Canada	*S.tuberosum*	1994
PP026B_184_111*	GQ853606	South Africa	*S.tuberosum*	
PRI-509	EU563512	The Netherlands	*S.tuberosum*	1938
PVY-12*	AB185833	Syria	*S.tuberosum*	2003
PVY-C-CM	AJ390302	UK	*S.tuberosum*	
PVY-MN	AF463399	USA	*N.tabacum*	2001
PVY-NBR	AF255660	Brazil	*S.tuberosum*	
PVY-N-RB	AJ390285	UK	*S.tuberosum*	
PVYNTN1	GQ853632	South Africa	*S.tuberosum*	2007
PVYNTN17_1*	JN936429	South Africa	*S.tuberosum*	2007
PVYNTN3_3*	GQ853607	South Africa	*S.tuberosum*	
PVY-OBR	AF255659	Brazil	*S.tuberosum*	
PVY-Sumi*	EU885418	SouthKorea	*S.tuberosum*	2008
PVY-ThaiNguyen	FM201468	Vietnam	*S.tuberosum*	
RB	HM367076	Canada	*S.tuberosum*	
RRA-1	AY884984	USA	*S.tuberosum*	2001
S25774_1*	JQ954331	Switzerland	*S.tuberosum*	2008
S25776_3*	JQ954332	Switzerland	*S.tuberosum*	2008
S25777_4*	JQ954333	Switzerland	*S.tuberosum*	2008
S25781_8*	JQ954334	Switzerland	*S.tuberosum*	2008
S25783_10*	JQ954335	Switzerland	*S.tuberosum*	2008
S25789_16*	JQ954336	Switzerland	*S.tuberosum*	2008
S25907-134	JQ954393	Switzerland	*S.tuberosum*	2008
S25972-199*	JQ954394	Switzerland	*S.tuberosum*	2008
SASA-110	AJ585195	UK	*S.tuberosum*	1997
SASA-61	AJ585198	UK	*S.tuberosum*	1997
Satina*	AJ890347	Germany	*S.tuberosum*	2002
SC143	JQ954337	Scotland	*S.tuberosum*	1996
SC190	JQ954338	Scotland	*S.tuberosum*	2000

<u>SC61</u>	JQ954339	Scotland	*S.tuberosum*	1986
SCRI-N	AJ585197	UK	*S.tuberosum*	1985
SCRI-O	AJ585196	UK	*S.tuberosum*	1985
Shanxi	EU719650	China	*S.tuberosum*	
Si15_Italy	AJ303093	Italy	*C.annuum*	
Si15_Turkey	AJ303094	Turkey	*C.annuum*	
<u>SLO4</u>	JQ954376	Slovenia	*S.tuberosum*	2009
<u>SLO7*</u>	JQ954377	Slovenia	*S.tuberosum*	2007
S-NTN	AJ390295	UK	*S.tuberosum*	
SON41	AJ439544	France	*S.nigrum*	1972
S-RB96**	AJ390308	UK	*S.tuberosum*	
SS082A_171_4	GQ853601	South Africa	*S.tuberosum*	
SS082A_194_14*	GQ853608	South Africa	*S.tuberosum*	
SS082A_88	GQ853600	South Africa	*S.tuberosum*	2005
SS121_154_10	GQ853633	South Africa	*S.tuberosum*	
SS121_166_56	GQ853637	South Africa	*S.tuberosum*	
SS121_197_16*	GQ853610	South Africa	*S.tuberosum*	
SS121_53_42*	GQ853609	South Africa	*S.tuberosum*	
SS121_82_1*	GQ853612	South Africa	*S.tuberosum*	
SS147_144_144*	GQ853611	South Africa	*S.tuberosum*	
SYR-D4*	AB295477	Syria	*S.tuberosum*	2004
SYR-D9*	AB295478	Syria	*S.tuberosum*	
SYR-II-2-8	AB461451	Syria	*S.tuberosum*	2006
SYR-II-Be1	AB461452	Syria	*S.tuberosum*	2004
SYR-NB-16	AB270705	Syria	*N.tabacum*	2006
SYR-Sn	AB295475	Syria	*S.nigrum*	2004
T50	AB331544	Japan	*S.tuberosum*	
Tamarillo	FM244834	Taiwan	*C.betacea*	2008
<u>TC_2-186</u>	JQ954340	Czech Republic	*S.tuberosum*	2006
<u>TC_2-187</u>	JQ954341	Czech Republic	*S.tuberosum*	2006
<u>TC_2-191*</u>	JQ954343	Czech Republic	*S.tuberosum*	2006
<u>TC_2-196</u>	JQ954344	Czech Republic	*S.tuberosum*	2006
<u>TC_2-197</u>	JQ954345	Czech Republic	*S.tuberosum*	2006
<u>TC_2-198</u>	JQ954346	Czech Republic	*S.tuberosum*	2006
<u>TC_2-199</u>	JQ954347	Czech Republic	*S.tuberosum*	2006
<u>TC_2-200</u>	JQ954348	Czech Republic	*S.tuberosum*	2006
TT014_184_135*	GQ853613	South Africa	*S.tuberosum*	
TT019A_107_52	GQ853665	South Africa	*S.tuberosum*	
TT026B_195_58	GQ853602	South Africa	*S.tuberosum*	
TT026B_86_128	GQ853638	South Africa	*S.tuberosum*	
TT026B_88_115	GQ853639	South Africa	*S.tuberosum*	
TT138D_111_79	GQ853640	South Africa	*S.tuberosum*	
TT138D_13_68*	GQ853614	South Africa	*S.tuberosum*	
TT138E_102_96	GQ853641	South Africa	*S.tuberosum*	
TT138E_111_104	GQ853666	South Africa	*S.tuberosum*	
TT138E_113_106	GQ853642	South Africa	*S.tuberosum*	

TT141A_76_73	GQ853643	South Africa	*S.tuberosum*	
Tu_660	AY166866	Canada	*S.tuberosum*	1991
Tu12.3	AJ303095	Turkey	*C.annuum*	
TU619	AJ390309	USA	*S.tuberosum*	
US04_24	JQ954392	USA	*S.tuberosum*	2004
US05_11	JQ954373	USA	*S.tuberosum*	2005
US05_12*	JQ954391	USA	*S.tuberosum*	2005
US05_13	JQ954349	USA	*S.tuberosum*	2005
US05_14	JQ954350	USA	*S.tuberosum*	2005
US05_17	JQ954385	USA	*S.tuberosum*	2005
US05_19	JQ954351	USA	*S.tuberosum*	2005
US05_20	JQ954386	USA	*S.tuberosum*	2005
US05_25	JQ954352	USA	*S.tuberosum*	2005
US05_26	JQ954353	USA	*S.tuberosum*	2005
US05_28	JQ954389	USA	*S.tuberosum*	2005
US05_3	JQ954354	USA	*S.tuberosum*	2005
US05_30	JQ954355	USA	*S.tuberosum*	2005
US05_31	JQ954390	USA	*S.tuberosum*	2005
US05_33	JQ954356	USA	*S.tuberosum*	2005
US05_36	JQ954388	USA	*S.tuberosum*	2005
US05_37	JQ954375	USA	*S.tuberosum*	2005
US05_39	JQ954357	USA	*S.tuberosum*	2005
US05_41	JQ954358	USA	*S.tuberosum*	2005
US05_45	JQ954359	USA	*S.tuberosum*	2005
US05_48	JQ954360	USA	*S.tuberosum*	2005
US05_49	JQ954361	USA	*S.tuberosum*	2005
US05_51	JQ954362	USA	*S.tuberosum*	2005
US05_52*	JQ954363	USA	*S.tuberosum*	2005
US05_56	JQ954364	USA	*S.tuberosum*	2005
US05_6	JQ954365	USA	*S.tuberosum*	2005
US05_64	JQ954366	USA	*S.tuberosum*	2005
US05_7	JQ954367	USA	*S.tuberosum*	2005
US05_9	JQ954368	USA	*S.tuberosum*	2005
US06_52	JQ954369	USA	*S.tuberosum*	2006
US06_55	JQ954370	USA	*S.tuberosum*	2006
US06_56	JQ954371	USA	*S.tuberosum*	2005
US06_59	JQ954372	USA	*S.tuberosum*	2005
USMN20	JQ954374	USA	*S.tuberosum*	2004
v942490*	EF016294	UK	*S.tuberosum*	1994
v951156-1	AJ390286	UK	*S.tuberosum*	
v951175	AJ390304	UK	*S.tuberosum*	
v951204	AJ390292	UK	*S.tuberosum*	
v951204-N**	AJ390291	UK	*S.tuberosum*	
v951218	AJ390287	UK	*S.tuberosum*	
v97005	AJ390303	UK	*S.tuberosum*	
Wilga	EF558545	Poland	*S.tuberosum*	1984

WW002_22_147*	GQ853615	South Africa	*S.tuberosum*	
WW002_74_150*	GQ853616	South Africa	*S.tuberosum*	
WW002_82_151*	GQ853617	South Africa	*S.tuberosum*	
WW010_146_164	GQ853645	South Africa	*S.tuberosum*	
WW010_147_166	GQ853646	South Africa	*S.tuberosum*	
WW010_70_158	GQ853644	South Africa	*S.tuberosum*	
WW154_175_62	GQ853647	South Africa	*S.tuberosum*	
WW154A_62_86	GQ853648	South Africa	*S.tuberosum*	
WW202B_21_172*	GQ853618	South Africa	*S.tuberosum*	
WW202B_24_184*	GQ853619	South Africa	*S.tuberosum*	
WW282E_3	GQ853649	South Africa	*S.tuberosum*	
Xinyang	EU719648	China	*S.tuberosum*	
Z14*	JN936440	South Africa	*S.tuberosum*	2009
Z16*	JN936441	South Africa	*S.tuberosum*	2010
Z26*	GQ853620	South Africa	*S.tuberosum*	2005
Zhuanglang103	EF592523	China	*N.tabacum*	
	GQ496607	Latvia	*S.tuberosum*	2007
	AM931253	China	*N.tabacum*	
	AM931254*	China	*S.tuberosum*	

*isolates showing the common recombination point at position 9170
**isolates showing other recombination points at the CP cistron
Underlined isolates in column one, newly described in this paper.

Table S2. Amino acid composition for potato (P) and non-potato (NP) isolates (180 and 32, respectively) at positively selected codons. The last two columns indicate those amino acids that have been detected only in P or NP isolates, respectively, for a given position. Codon positions are given as the corresponding amino acid positions in the CP cistron.

Position	**P composition**	**NP composition**	**P specific**	**NP specific**
1	113A, 60G, 5V	24A, 5G, 3V		
68	173E, 4K, 3G	31E, 1D	K, G	D
193	98V, 57G, 25M	11G, 8V, 7I, 4M, 1R, 1T		I, R, T
216	178A, 1E, 1G	32A	E, G	

Table S3. Amino acid composition for potato (P) and non-potato (NP) isolates (180 and 32, respectively) at covarying codons. The last two columns indicate those amino acids that have been detected only in potato or non-potato isolates, respectively, for a given position. Codon positions are given as the corresponding amino acid positions in the CP cistron.

Position	P composition	NP composition	P specific	NP specific
1	113A, 60G, 5V	24A, 5G, 3V		
11	113S, 45T, 21N, 1A	22S, 9N, 1T	A	
17	125P, 53Q, 1L, 1R	28P, 4Q	L, R	
24	116S, 59P, 3L, 2R	16P, 9S, 5R, 2L		
26	119P, 61L,	27P, 4L, 1S		S
29	126G, 54E	27G, 4E, 1A		A
31	126D, 53E, 1V	25D, 7E	V	
99	122M, 53L, 5V	17M, 8V, 5L, 1T, 1I		T, I
138	94D, 86N	14S, 10N, 8D		S

Table S4. Amino acid composition for PVY^C, PVY^O and PVY^N strain isolates (22, 132 and 57 isolates, respectively) at positively selected codons. The last three columns indicate those amino acids that have been detected only in PVY^C, PVY^O or PVY^N groups, respectively, for a given position. Codon positions are given as the corresponding amino acid positions in the CP cistron.

Position	**PVY^C composition**	**PVY^O composition**	**PVY^N composition**	**PVY^C specific**	**PVY^O specific**	**PVY^N specific**
1	20A, 2V	108A, 18G, 4V	47G, 8A, 2V			
68	21E, 1D	132E	50E, 4K, 3G	D		K, G
193	13G, 6I, 1M, 1T, 1V	104V, 28M	55G, 1R, 1V	I, T		R
216	22A	130A, 1E, 1G	57A		E, G	

Table S5. Amino acid composition for PVY^C, PVY^O and PVY^N strain isolates (22, 132 and 57 isolates, respectively) at covarying codons. The last three columns indicate those amino acids that have been detected only in PVY^C, PVY^O or PVY^N groups, respectively, for a given position. Codon positions are given as the corresponding amino acid positions in the CP cistron.

Position	**PVY^C composition**	**PVY^O composition**	**PVY^N composition**	**PVY^C specific**	**PVY^O specific**	**PVY^N specific**
1	20A, 2V	108A, 18G, 4V	47G, 8A, 2V			
11	19S, 3N	114S, 18N	46T, 9N, 1S, 1A			T, A
17	22P	130P, 1Q, 1L	56Q, 1R		L	R
24	11P, 7R, 3S, 1L	122S, 7P, 3L	56P, 1L	R		
26	21P, 1S	122P, 10L	55L, 2P	S		
29	21G, 1A	131G, 1E	57E	A		
31	19D, 3E	131D, 1V	57E		V	
99	12V, 7M, 1I, 1L, 1T	132M	57L	V, I, T		
138	14S, 6N, 2D	79N, 53D	47D, 10N	S		

Table S6. Amino acid composition for PVY^C, PVY^O and PVY^N strain isolates (22, 132 and 57 isolates, respectively) at the region showing evidence of positive selection for the internal branch leading to PVY^C clade (branch b2, codons 187-194, shown in Figure S1). The last three columns indicate those amino acids that have been detected only in PVY^N, PVY^O or PVY^C groups, respectively, for a given position.

Position	**PVY^C composition**	**PVY^O composition**	**PVY^N composition**	**PVY^C specific**	**PVY^O specific**	**PVY^N specific**
187	19I, 1H, 1N, 1V	130I, 2T	55V, 2I	H, N	T	
188	22R	132R	57R			
189	22N	132N	56N, 1T			T
190	22L	132L	56L, 1V			V
191	22R	132R	57R			
192	21D, 1V	132D	57D	V		
193	13G, 6I, 1M, 1T, 1V	104V, 28M	55G, 1R, 1V	I, T		R
194	17S, 5G	132G	57S			

Figure S1. MCC phylogeny of the PVY isolates for the CP cistron. The tree was calculated from the posterior distribution of trees generated by Bayesian MCMC coalescent analyses with BEAST (Drummond & Rambaut, 2007). Posterior probabilities are indicated above branches. Branches detected to be under positive selection are shown in red, and internal branches are identified numbering in the range b1-b18. For clarity, branches were transformed as proportional using FigTree (www.tree.bio.ed.ac.uk).