University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Papers in Plant Pathology | Plant Pathology Department

1-1-2005

Genome sequences of *Agropyron mosaic virus* and *Hordeum mosaic virus* support reciprocal monophyly of the genera *Potyvirus* and *Rymovirus* in the family *Potyviridae*

Roy C. French
University of Nebraska - Lincoln, rfrench2@unl.edu

D.C. Stenger
University of Nebraska - Lincoln

Follow this and additional works at: http://digitalcommons.unl.edu/plantpathpapers

Part of the Plant Pathology Commons

French, Roy C. and Stenger, D.C., "Genome sequences of *Agropyron mosaic virus* and *Hordeum mosaic virus* support reciprocal monophyly of the genera *Potyvirus* and *Rymovirus* in the family *Potyviridae*" (2005). *Papers in Plant Pathology.* Paper 7.
http://digitalcommons.unl.edu/plantpathpapers/7

This Article is brought to you for free and open access by the Plant Pathology Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in Plant Pathology by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Arch Virol (2005) 150: 299–312
DOI 10.1007/s00705-004-0396-6

Archives of Virology
Printed in Austria

Genome sequences of *Agropyron mosaic virus* and *Hordeum mosaic virus* support reciprocal monophyly of the genera *Potyvirus* and *Rymovirus* in the family *Potyviridae*

R. French and **D. C. Stenger**

Agricultural Research Service, United States Department of Agriculture
and Department of Plant Pathology, University of Nebraska,
Lincoln, Nebraska, U.S.A.

Received March 19, 2004; accepted July 21, 2004
Published online September 24, 2004 © Springer-Verlag 2004

Summary. Assignment of mite-transmitted species to the genus *Rymovirus* (family *Potyviridae*) has changed several times, and the status of the genus has been questioned. To address this issue, complete genome sequences of the rymoviruses *Agropyron mosaic virus* (AgMV) and *Hordeum mosaic virus* (HoMV) were determined. AgMV (9540 nucleotides) and HoMV (9463 nucleotides) each encode a single polyprotein with proteinase cleavage sites demarcating protein products characteristic of monopartite species of the family *Potyviridae*. Of the described species of *Potyviridae*, AgMV and HoMV are most closely related to each other (68.5% nucleotide and 71.6% amino acid sequence identity) and equidistant (about 53% nucleotide and about 49% amino acid sequence identity) from a third rymovirus, *Ryegrass mosaic virus* (RGMV). Phylogenetic analyses by neighbor joining, maximum parsimony, and Bayesian inference each grouped the three *Rymovirus* species in an exclusive clade distinct from a clade containing 34 species of the genus *Potyvirus*. Because AgMV, HoMV, and RGMV share a reciprocal monophyletic relationship with species of the genus *Potyvirus* and are divergent in sequence and type of vector, the genus *Rymovirus* should be retained as a taxonomic unit within the family *Potyviridae*.

Introduction

The genus *Rymovirus* was established to distinguish species of the family *Potyviridae* that were known or suspected to be transmitted by eriophyid mites from aphid-transmitted species of the genus *Potyvirus* [38, 41]. As originally constituted, the genus *Rymovirus* included as definitive or tentative members

Ryegrass mosaic virus (RGMV), *Agropyron mosaic virus* (AgMV), *Hordeum mosaic virus* (HoMV), *Wheat streak mosaic virus* (WSMV), *Brome streak mosaic virus* (BrSMV), *Oat necrotic mottle virus* (ONMV), and Spartina mottle virus (SpMV). RGMV was designated as the type species of the genus, even though at that time partial sequence was available only for WSMV [22]. However, as sequences became available for other species, it became evident that the genus *Rymovirus* was a paraphyletic assemblage of at least two distinct lineages [14, 28, 36].

Presently, complete nucleotide (nt) sequences are known for two strains of RGMV [21, 31], five strains of WSMV [3, 24, 36], and one strain each of BrSMV [12] and ONMV [35]. Only partial sequences representing about 2 kb of the 3′-proximal region of the genomes of AgMV and HoMV have been determined [27]. A recent phylogenetic treatment of these viruses [24] indicates that WSMV, BrSMV, and ONMV share a common ancestor and constitute a distinct clade distantly related to the whitefly-transmitted *Sweet potato mild mottle virus* (SPMMV) of the genus *Ipomovirus*. As a result, WSMV, BrSMV, and ONMV have been reassigned to the genus *Tritimovirus*, with WSMV designated as the type species. Analysis of the CP sequence of SpMV [11] suggests that this species is not closely related to either rymoviruses or tritimoviruses, and may warrant placement in a new monotypic genus within the family *Potyviridae*. Thus, the genus *Rymovirus* currently retains only three species (RGMV, AgMV, and HoMV) that, based on 3′-proximal sequences, share a most recent common ancestor with species of the genus *Potyvirus* [14, 24, 28, 33, 36].

Despite differences in vector taxa distinguishing RGMV, AgMV, and HoMV from aphid-transmitted potyviruses, common ancestry based on CP amino acid (aa) sequences have caused some to question whether the genus *Rymovirus* should remain a taxon distinct from the genus *Potyvirus* [33]. Thus, the status of the genus *Rymovirus* is in doubt. To resolve this issue, we describe the cloning and sequencing of the complete genomes of the rymoviruses AgMV and HoMV, and through distance and phylogenetic analyses re-evaluate the genus *Rymovirus* as a taxonomic unit within the family *Potyviridae*.

Materials and methods

Cloning and sequencing of AgMV and HoMV

The AgMV isolate (ND402) used in this study represents a "mild green mosaic" strain from *Agropyron repens* (L.) Beauv. obtained from the collection of R. G. Timian. The HoMV isolate (ATCC PV81) used in this study is the same isolate for which partial sequence was determined [27] and was originally isolated from diseased barley grown in Alberta in the 1960s [34]. Both viruses were propagated by mechanical inoculation of wheat (*Triticum aestivum* L.) cv. Arapahoe. The identity of each virus isolate was authenticated by sequencing of reverse transcription-polymerase chain reaction products (data not shown) using primers specific for AgMV or HoMV that were based on the 3′-proximal partial sequences of AgMV (U30615) and HoMV (U30616) previously reported by Salm et al. [27].

Virion purification, virion RNA extraction, and cDNA cloning of the AgMV and HoMV genomes were accomplished using essentially the same methods as described for ONMV

[35]. Multiple overlapping cDNA clones were completely sequenced on both strands by automated sequencing (Davis Sequencing, Inc., Davis, CA, and DNA Sequencing Facility, Iowa State University) using custom and universal primers. Sequences of individual clones were compiled using Sequencher 4.1 (Gene Codes, Ann Arbor) to generate a consensus sequence for each viral genome in which each nt position was determined from a minimum of two independent clones, or more in cases where a nt position exhibited polymorphism.

Sequence analyses

Sequence comparisons were conducted with taxa of the genera *Rymovirus* (four sequences representing three species), *Tritimovirus* (three species), *Ipomovirus* (one species), and *Potyvirus* (34 species) for which complete nt sequences have been determined. The virus species (with GenBank accession numbers in parentheses) included in the analyses were for the genus *Rymovirus*, AgMV (AY623626), HoMV (AY623627), RGMV strain DA (RGMV-DA, NC_001814), RGMV strain AV (RGMV-AV, AF035818); for the genus *Tritimovirus*, BrSMV (NC_003501), ONMV (AY377938), and WSMV (NC_001886); for the genus *Ipomovirus*, SPMMV (NC_003797); and for the genus *Potyvirus*, *Bean common mosaic virus* (BCMV, NC_003397), *Bean common mosaic necrosis virus* (BCMNV, NC_004047), *Bean yellow mosaic virus* (BYMV, NC_003492), *Clover yellow vein virus* (ClYVV, NC_003536), *Cocksfoot streak virus* (CSV, NC_003742), *Cowpea aphid-borne mosaic virus* (CABMV, NC_004013), *Dasheen mosaic virus* (DsMV, NC_003537), *Japanese yam mosaic virus* (JYMV, NC_000947), *Johnsongrass mosaic virus* (JGMV, NC_003606), *Leek yellow stripe virus* (LYSV, NC_004011), *Lettuce mosaic virus* (LMV, NC_003605), *Maize dwarf mosaic virus* (MDMV, NC_003377), *Onion yellow dwarf virus* (OYDV, NC_005029), *Papaya leaf-distortion mosaic virus* (PLDMV, NC_005028), *Papaya ringspot virus* (PRSV, NC_001785), *Pea seed-borne mosaic virus* (PSBMV, NC_001671), *Peanut mottle virus* (PeMoV, NC_002600), *Pepper mottle virus* (PepMoV, NC_001517), *Peru tomato mosaic virus* (PToMV, NC_004573), *Plum pox virus* (PPV, NC_001445), *Potato virus A* (PVA, NC_004039), *Potato virus V* (PVV, NC_004010), *Potato virus Y* (PVY, NC_001616), *Scallion mosaic virus* (ScaMoV, NC_003399), *Sorghum mosaic virus* (SrMV, NC_004035), *Soybean mosaic virus* (SMV, NC_002634), *Sugarcane mosaic virus* (SCMV, NC_003398), *Sweet potato feathery mottle virus* (SPFMV, NC_001841), *Tobacco etch virus* (TEV, NC_001555), *Tobacco vein mottling virus* (TVMV, NC_001768), *Turnip mosaic virus* (TuMV, NC_002509), *Wild potato mosaic virus* (WPMV, NC_004426), *Yam mosaic virus* (YMV, NC_004752), and *Zucchini yellow mosaic virus* (ZYMV, NC_0033224).

Deduced aa sequences were aligned with ProAlign [20] using the Blossum62 similarity matrix [16]. This alignment was converted into an alignment of nt sequences with tranalign, part of the EMBOSS package [25]. Uniformity of aa or nt composition among sequences was tested using TREE-PUZZLE [30]. TREE-PUZZLE also was used to estimate values of alpha (the shape parameter for approximating among-site rate heterogeneity as a gamma distribution) for the aa sequence alignments. Phylogenetic trees were reconstructed by the neighbor joining (NJ) [26] method from CI and NIb aa distances (calculated using a simple identity matrix, with corrections for among-site rate heterogeneity) as implemented in MEGA2 [19]. Consensus trees were generated from 500 bootstrap replicates. Since bootstrap values equal to or greater than 70% have been shown empirically to provide reliable support for phylogenetic groupings [17], a majority rule bootstrap cutoff of 70% was used to construct consensus trees. Bayesian phylogenetic inferences were made using MrBayes version 3.0 [18], starting with a random tree and employing the HKY model of DNA substitution [15]. Four Markov Chain Monte Carlo (MCMC) analyses were run for 100,000 generations, sampling every 100th generation. The process was repeated twice with different random starting trees to confirm

that independent runs of MCMC analyses were converging on the same parameter values and phylogenetic trees. Maximum parsimony (MP) and NJ [BIONJ method, 8] analyses, with 500 bootstrap replicates each, were done with PAUP* version 4b10 [37]. To examine whether there was a conflict in the phylogenetic information content between the CI or NIb aa alignment data, the incongruence length difference test (ILD) [6] was conducted using PAUP (in which it is called the Partition Homogeneity Test), with 1000 permutations. Parsimony uninformative sites were removed prior to this analysis. The effect of noise due to homoplasy on the ILD test was determined as described by Dolphin et al. [4] with ten replicates of shuffled data each for the CI and NIb alignments. Characters in each column of the alignments were randomized using the "shuffle states among taxa" module of the computer program Mesquite (W. P. Maddison, D. R. Maddisson, Mesquite: a modular system for evolutionary analysis, version 1.0, University of Arizona, 2003). Shimodaira–Hasegawa tests [10, 32] were conducted using PAUP*, with resampling estimated log-likelihood (RELL) optimization and 10,000 bootstrap replicates. Majority-rule consensus phylograms were visualized using TREEVIEW 1.5.3 [23] with the ipomovirus SPMMV designated as the outgroup.

Results and discussion

Annotation and comparison of Rymovirus sequences

The complete nt sequences (exclusive of the variable-length polyadenylated tail) of AgMV and HoMV were 9540 nt and 9463 nt, respectively. Comparison of the complete sequences reported here with previously determined partial sequences [27] confirmed the identity of each virus examined. The AgMV sequence shared 97.5% nt sequence identity with the partial sequence U30615 that was derived from a different isolate of AgMV, whereas the HoMV sequence shared 99.2% nt sequence identity with the partial sequence U30616 derived from this same isolate of HoMV.

The genomes of AgMV and HoMV each encode a polyprotein with an initiation codon located at nts 132–134, in a genomic position similar to that of RGMV-AV (nts 111–113) or RGMV-DA (nts 113–115), and a termination codon at nts 9366–9368 (AgMV) or 9282–9284 (HoMV). AgMV and HoMV also have initiation codons (nts 6–8) upstream of and in-frame with the polyprotein open reading frame (ORF) but this ORF terminates at nt 56 in HoMV. An additional in-frame initiation codon (nts 21–23) is present in AgMV but not HoMV. Nonetheless, the conserved AUG at nts 132–134 was considered as the most likely site of polyprotein translation initiation in both viruses, as members of the family *Potyviridae* typically have 5′-untranslated regions (5′-UTRs) of more than 100 nts. The AgMV and HoMV polyproteins are likely cleaved by the viral proteinases P1 [39], HC-Pro [2], and NIa [5] to generate protein products (P1, HC-Pro, P3, 6K1, CI, 6K2, NIa-VPg, NIa-Pro, NIb, and CP) typical of monopartite species of the family *Potyviridae*. Comparison of polyprotein proteinase cleavage sites deduced for all three rymovirus species indicated that fewer substitutions were present when comparing AgMV to HoMV, relative to pairwise comparisons of RGMV with the other rymovirus species (Table 1). For those sites cleaved by NIa proteinase, a valine residue was always present in the minus 4 position, and

Table 1. Predicted *Rymovirus* polyprotein proteinase cleavage sites

Cleavage site	Proteinase	RGMV-AV[a]	AgMV[b]	HoMV[b]
P1/HC-Pro	P1	HTILHY/S[c]	KRIQHF/S	NAIEHF/S
HC-Pro/P3	HC-Pro	LDYNVG/G	AEYNVG/G	KEYNVG/G
P3/6K1	NIa	DTVTHQ/A	EEVEHQ/A	EEVKHQ/A
6K1/CI	NIa	TVVTHQ/S	DAVYHQ/S	GSVYHQ/S
CI/6K2	NIa	TLVHHQ/S	TLVEHQ/S	ALVEHQ/S
6K2/NIa	NIa	TRVHLE/G	EQVTFE/G	EQVSFE/G
NIa-VPg/NIa-Pro	NIa	YSVQHE/S	SCVAHE/S	SQVGHE/S
NIa/NIb	NIa	EAVSHQ/S	TDVEHQ/H	TDVEHQ/H
NIb/CP	NIa	TKVVHE/A[d]	TLVYHE/A	TLVFHE/A

[a]Identical for RGMV-DA unless otherwise noted
[b]Underline denotes amino acid residues different from RGMV
[c]RGMV-DA cleavage site is HTIKHY/S
[d]Listed as SSAATQ/T in GenBank accession AF035818 for RGMV-AV

a histidine residue was present in the minus 2 position except for the 6K2/NIa cleavage site in which leucine (RGMV) or phenyalanine (AgMV and HoMV) was substituted for histidine.

Complete genome comparisons revealed that AgMV and HoMV were more closely related to each other (68.5% nt sequence identity, 71.6 aa% sequence identity) than to RGMV (ca. 53% nt, ca. 49% aa). Among individual cistrons, highest sequence identity between AgMV and HoMV occurred within 6K1 (78.6% nt, 84.9% aa) whereas the P1 cistron was most divergent (51.7% nt, 46.5% aa). The NIb cistrons of AgMV and HoMV had the highest aa identities (ca. 65–66%) relative to RGMV.

Phylogenetic relationships within the family Potyviridae

The deduced amino acid sequences of AgMV and HoMV were aligned with those of 40 other members of the *Potyviridae* on a cistron by cistron basis with a probabilistic, hidden Markov model algorithm for alignment employed by the ProAlign computer program. In addition to the matrix of aligned sequences, this program provides a method to assess how ambiguous each column in the alignment might be by resampling the posterior frequencies of alternative local alignments for each cell in the alignment matrix. The score for the most ambiguous cell is summarized as the minimum posterior probability of an alignment column, and these are plotted for HC-Pro, CI, NIa, NIb, and CP (Fig. 1). Computer memory was insufficient to allow resampling of the more divergent P1 and P3 alignments; consequently these sequences were not used in subsequent analyses.

Uncertainty in sequence alignment is not the only factor that may confound phylogenetic analysis of a data set [13, 29]. One of the better known artifacts of MP is long branch attraction where there is a tendency for more divergent taxa to

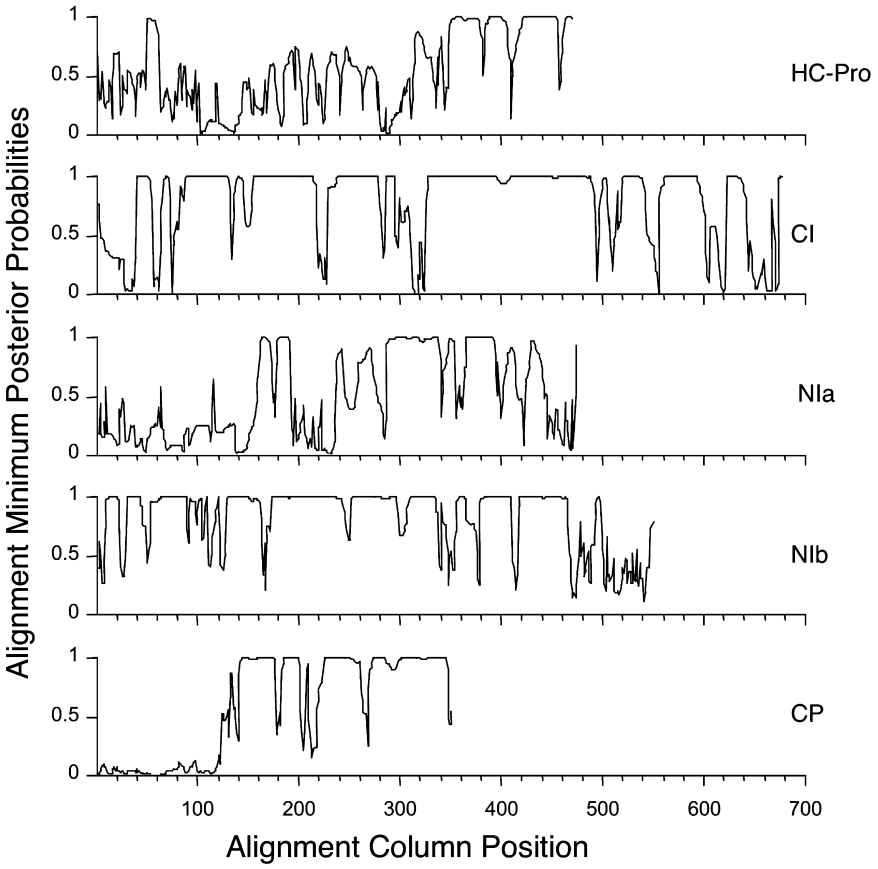

Fig. 1. Confidence estimations for amino acid sequence alignments of five *Potyviridae* cistrons (HC-Pro, CI, NIa, NIb, and CP). Presented are alignment minimum posterior probabilities plotted as a function of alignment column position generated by the ProAlign computer program [20]. The y axis is essentially the lowest probability of the alignment, given the data, and thus provides a relative measure the reliability of the alignment

be grouped together [7]. Distance methods also do poorly when among-sequence diversity is high. Average sequence identity between species was 53% and ranged from 77% to 40%. Uneven nt or aa composition among sequences also can bias phylogenetic inference. Chi-square tests for uniformity in nt composition were rejected for 29 of the 42 taxa. Since these characteristics of the data suggest potential difficulties in phylogenetic reconstruction, several approaches were taken. NJ of aa distances, corrected for among-site rate heterogeneity was done for CI and NIb, the two alignments with the fewest gaps. The concatenated data set (HC-Pro, CI, NIb, and CP) was analyzed using NJ, MP, and a likelihood-based Bayesian analysis (BA).

No taxa had skewed amino acid composition for either CI or NIb. For CI, the average pairwise aa sequence identities were 71% among the *Rymovirus* taxa, 53% among *Potyvirus* taxa, and a mean identity between the two genera was 48%. NIb had less divergence, with average aa sequence identities among

taxa of 74% (*Rymovirus*) and 57% (*Potyvirus*), and a mean identity between the two genera of 55%. The estimated gamma-shape parameter was 0.7 for CI (strong rate heterogeneity) and 1.2 for NIb (weak rate heterogeneity). Seventy percent majority rule (500 bootstrap replications) consensus trees were produced by the NJ method for CI and NIb. The CI tree (Fig. 2) lacked phylogenetic resolution,

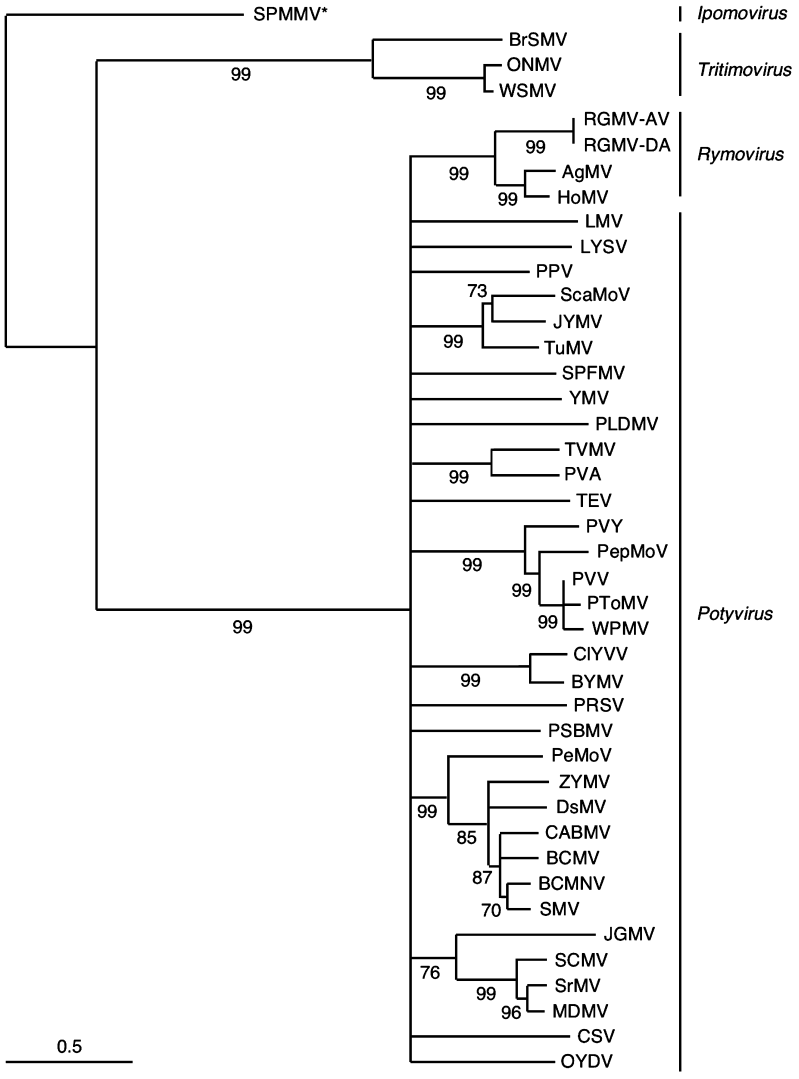

Fig. 2. Phylogenetic relationships among CI proteins encoded by species of the family *Potyviridae*. Presented is a neighbor-joining phylogram (500 bootstrap replications) based on an alignment of CI amino acid sequences. Numbers along branches indicate bootstrap support rounded to nearest percentage. Only nodes having ≥70% bootstrap support are shown, all others were collapsed to polytomies. Horizontal branch lengths are proportional to genetic distance; the scale bar indicates a branch length corresponding to 0.5 substitutions per amino acid position. Asterisk denotes taxon designated as the outgroup used to root the phylogram. The generic affiliations of virus taxa are indicated on the right

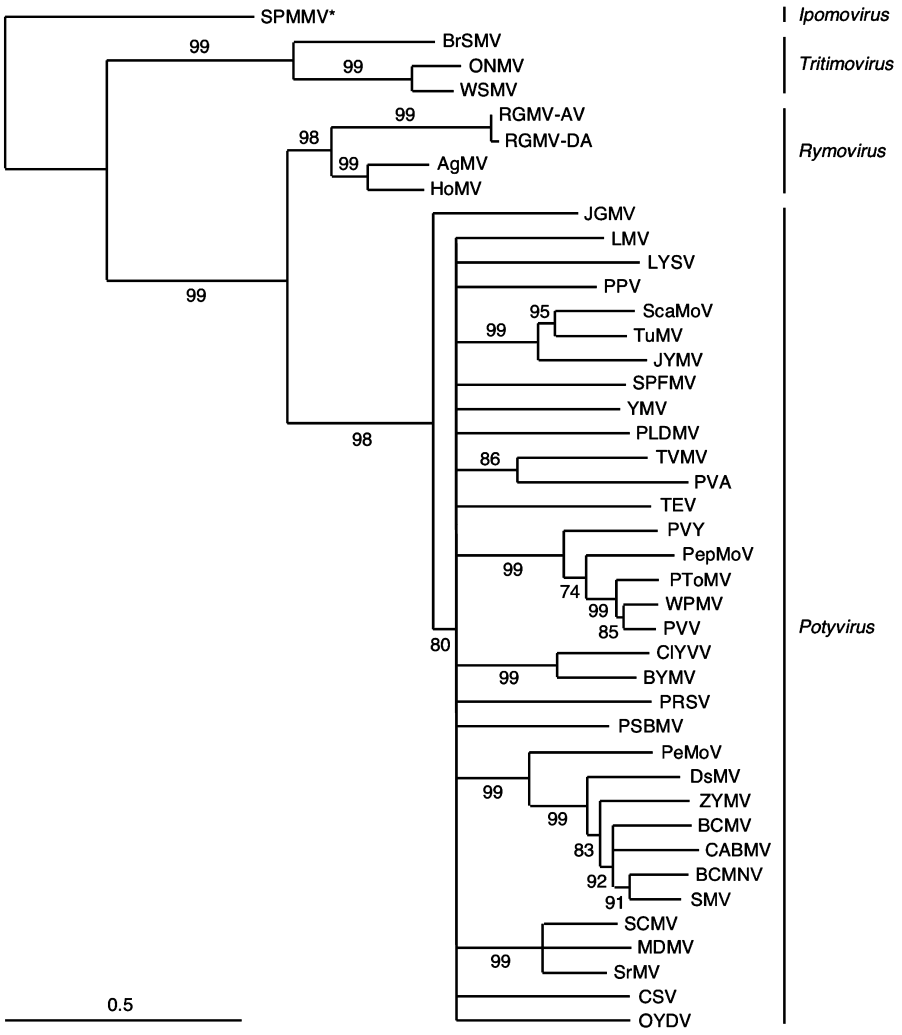

Fig. 3. Phylogenetic relationships among NIb proteins encoded by species of the family *Potyviridae*. Presented is a neighbor-joining phylogram (500 bootstrap replications) based on an alignment of NIb amino acid sequences. Numbers along branches indicate bootstrap support rounded to nearest percentage. Only nodes having ≥70% bootstrap support are shown, all others were collapsed to polytomies. Horizontal branch lengths are proportional to genetic distance; the scale bar indicates a branch length corresponding to 0.5 substitutions per amino acid position. Asterisk denotes taxon designated as the outgroup used to root the phylogram. The generic affiliations of virus taxa are indicated on the right

with *Rymovirus* species forming a single clade as part of a polytomy with all *Potyvirus* taxa. In contrast, *Rymovirus* and *Potyvirus* taxa grouped separately as monophyletic sister clades in the NIb tree (Fig. 3).

Even though the CI and NIb alignments had the fewest gaps, the two cistrons still may be evolving at different rates and possibly may have different phylogenetic histories. As noted above, the amount of site-to-site substitution rate heterogeneity was higher for CI than NIb. To further test for differences between

the two cistrons, the ILD test was used. This is a permutation-based statistical method to determine similarity of phylogenetic signal between two sets of data. The null hypothesis of no differences was rejected by this test ($P = 0.034$) suggesting that CI and NIb have either different histories, were subject to different selective constraints, differ in substitution saturation due to multiple replacements at some sites, or a combination of all three possibilities. To test whether the ILD was larger than might be expected from that due to homoplasy, the test was repeated by separately shuffling characters in the alignment matrix of either CI or NIb. The shuffled alignments, therefore, serve as data sets approximating pure noise. For unshuffled data, the ILD was 65, while for CI shuffled data the ILD was 262 ± 21 (mean of 10 replicates with standard deviation) and 274 ± 34 for NIb shuffled data. An ILD for two sequences strongly differing in phylogenetic information should exceed that due to noise only. Thus, difference in phylogenetic information between the CI and NIb data most likely resulted from difference in noise level due to homoplasy rather than conflicting phylogenetic signals.

Since separate analysis of NIb and CI led to different, albeit not entirely inconsistent, inferences, HC-Pro, CI, NIa, NIb, and CP sequences were concatenated and subjected to several additional analyses: MP, NJ, and likelihood-based BA. The data were first examined for several systematic biases which could affect phylogenetic reconstruction. ONMV, RGMV-AV, and RGMV-DA had aa frequencies that rejected ($P < 0.05$) the null hypothesis of no differences from the overall average by a Chi-square test. Similar tests for similarities in nucleotide composition were rejected for 29 of the 42 taxa when all three codon positions were included, but failed for only BrSMV, AgMV, HoMV, RGMV-AV, RGMV-DA, CSV, OYMV, PLDMV, ClYVV, BYMV, and PSBMV when third codon position sites were excluded from the analysis. There was evidence of among-site rate heterogeneity for both aa (gamma-shape parameter, 0.67) and nt (gamma-shape parameter, 0.61) sequence data.

MP and NJ analyses were done using aa sequence data and 70% majority rule consensus trees were constructed for each (500 bootstrap replications). BA was done using nt sequence data excluding third codon position data. MCMC analyses were run for 100,000 generations, sampling every 100th generation. Tree likelihoods became stable after 20,000 generations and the last 500 trees (generations 50,100 to 100,000) were used to generate a 95% majority rule consensus tree. Finally, MP, NJ, and BA consensus trees were used to produce a strict consensus tree (Fig. 4). NJ and MP bootstrap values, and BA clade probabilities are given for each surviving node. The homoplasy index score for aa data on this strict consensus tree was 0.515. The strict consensus tree (Fig. 4) was similar to the NIb NJ tree (Fig. 3) except for the relative placement of OYMV and JGMV among the other potyviruses and the increased number of nodes with $\geq$70% bootstrap support. Thus, while there was uncertainty in sequence alignments, use of the larger concatenated data allowed inference of a more robust consensus tree.

Finally, the Shimodaira–Hasegawa test was used to statistically compare the three alternative phylogenetic hypotheses represented in Figs. 2–4 using

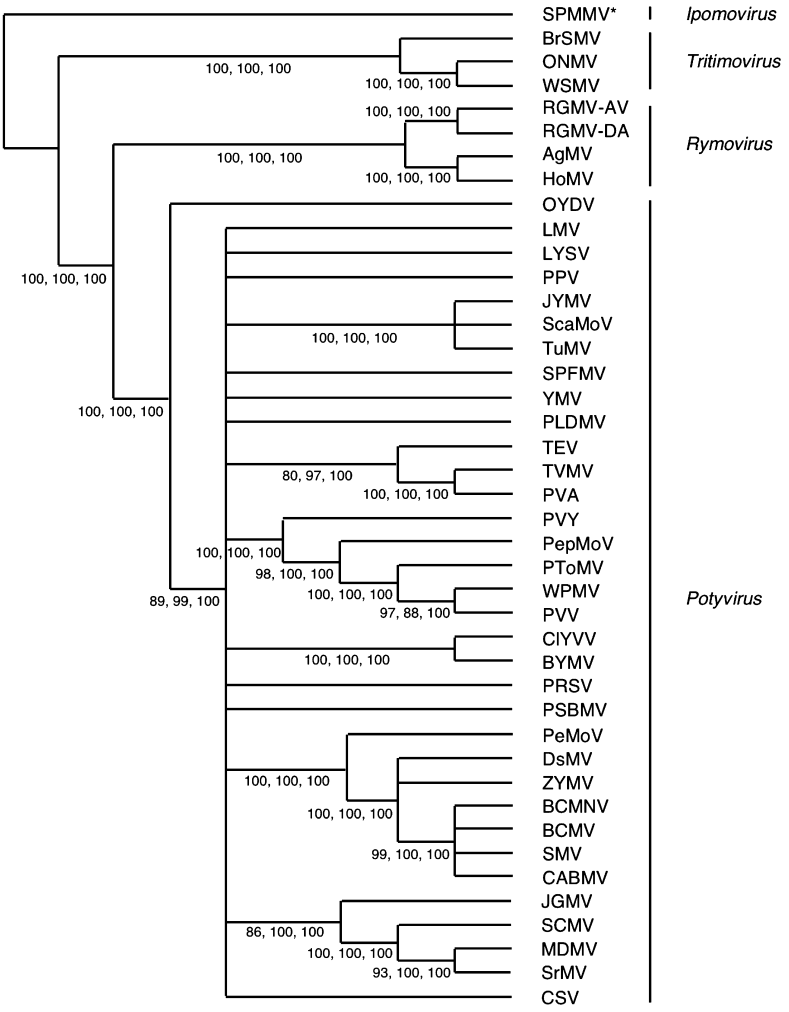

Fig. 4. Strict consensus phylogenetic tree of Neighbor Joining and Maximum Parsimony analysis of concatenated (HC-Pro, CI, NIa, NIb, and CP) amino acid sequences, as well as Bayesian phylogenetic analysis of concatenated nucleotide sequences using first and second codon positions. Nodal bootstrap support values for Maximum Parsimony (first number) and Neighbor Joining (second number) are shown. The third number represents the posterior clade probability of each node by Bayesian analysis. Only nodes having $\geq$70% support by all three phylogenetic reconstruction methods are shown, with all others collapsed to polytomies. Trees were rooted using SPMMV as the outgroup. The generic affiliations of virus taxa are indicated on the right

nucleotide sequence alignments and the HKY substitution model parameters. The phylogenetic tree presented in Fig. 4, had the highest likelihood scores (lowest −ln likelihood values) (Table 2) using sequence data of CI alone, NIb alone, or the five concatenated cistrons, compared to the trees in Figs. 2 or 3. The Shimodaira–Hasegawa test probabilities also were highly significant for differences between trees. That CI and NIb did not favor their own trees may seem surprising, but note that these are not original NJ trees, but are majority rule consensus trees with

Table 2. Likelihoods and Shimodaira–Hasegawa test comparisons of three alternative phylogenetic trees using nucleotide sequence data

Phylogenetic hypothesis	CI cistron only			NIb cistron only			Concatenated cistrons (HC-Pro, CI, NIb, NIa, CP)		
	$-\ln$ lk[a]	Delta[b]	P[c]	$-\ln$ lk[a]	Delta[b]	P[c]	$-\ln$ lk[a]	Delta[b]	P[c]
CI (Fig. 2)	69956	182	$<10^{-4}$	52750	230	$<10^{-4}$	259455	740	$<10^{-4}$
NIb (Fig. 3)	69979	205	$<10^{-4}$	52657	137	$<10^{-4}$	259286	571	$<10^{-4}$
Concatenated cistrons (HC-Pro, CI, NIb, NIa, CP) (Fig. 4)	69774	–	–	52520	–	–	258715	–	–

[a]Tree likelihood score
[b]Difference between log likelihood for this tree and the best alternative tree
[c]Shimodaira–Hasegawa test probability of a null hypothesis that there is no difference between this tree and the best tree (10,000 bootstrap replications)

many nodes collapsed. This result implies that CI, NIb, and the concatenated data all contain consistent phylogenetic information.

Retention of the genus Rymovirus is warranted

Previous phylogenetic analysis of the family *Potyviridae* using 3′-proximal sequences have shown that terminal branches are long relative to internal branches with generally weak bootstrap support for dividing *Potyvirus* and *Rymovirus* taxa into independent clades [1, 9, 33]. Long branches are known to be problematic in phylogenetic reconstruction [7, 13, 29] and the data had other characteristics that may be expected to confound phylogenetic reconstruction such as varying degrees of alignment ambiguity, compositional heterogeneity, and large genetic distances among taxa [13, 29, 40]. Nevertheless, our phylogenetic analyses using three different methods (NJ, MP, and BA) and based on a concatenated sequence comprising about 70% of the genome consistently placed the three known species of the genus *Rymovirus* in an exclusive clade distinct from a sister clade containing 34 aphid-transmitted species of the genus *Potyvirus*. Thus, the hypothesis that *Rymovirus* and *Potyvirus* comprise reciprocal monophyletic clades has robust support. Because AgMV, HoMV, and RGMV share a reciprocal monophyletic relationship with species of the genus *Potyvirus*, and are mutually divergent in sequence and type of vector, we conclude that the genus *Rymovirus* should be retained as a taxonomic unit within the family *Potyviridae*.

Acknowledgements

We thank B. Young and J. S. Hall for technical assistance. Mention of proprietary or brand names are necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standard of the product, and the use of the name by USDA implies no approval to the exclusion of others that also may be suitable. This article is in the public domain and not copyrightable. It may be freely reprinted with customary crediting of source.

References

1. Berger PH, Wyatt SD, Shiel PJ, Silbernagel MJ, Druffel K, Mink GI (1997) Phylogenetic analysis of the Potyviridae with emphasis on legume-infecting potyviruses. Arch Virol 142: 1979–1999
2. Carrington JC, Cary SM, Parks TD, Dougherty WG (1989) A second proteinase encoded by a plant potyviral genome. EMBO J 8: 365–370
3. Choi I-R, Hall JS, Henry M, Zhang L, Hein GL, French R, Stenger DC (2001) Contributions of genetic drift and negative selection on the evolution of three strains of wheat streak mosaic tritimovirus. Arch Virol 146: 619–628
4. Dolphin K, Belshaw R, Orme CDL, Quicke DLJ (2000) Noise and incongruence: interpreting results of the incongruence length difference test. Mol Phylogenet Evol 17: 401–406
5. Dougherty WG, Cary SM, Parks TD (1989) Molecular genetic analysis of a plant virus polyprotein cleavage site: a model. Virology 171: 356–364
6. Farris JS, Källersjö M, Kluge AG, Bult C (1995) Constructing a significance test for incongruence. Syst Biol 44: 570–572
7. Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27: 401–410
8. Gascuel O (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol Biol Evol 14: 685–695
9. Glasa M, Kudela O, Subr Z (2003) Molecular analysis of the 3′ terminal region of the genome of Beet mosaic virus and its relation with other potyviruses. Arch Virol 148: 1863–1871
10. Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. Syst Biol 49: 652–670
11. Götz R, Huth W, Lesemann D-E, Maiss E (2002) Molecular and serological relationships of Spartina mottle virus (SpMV) strains from *Spartina* spec. and from *Cynodon dactylon* to other members of the *Potyviridae*. Arch Virol 147: 379–391
12. Götz R, Maiss E (1995) The complete nucleotide sequence and genome organization of the mite-transmitted brome streak mosaic rymovirus in comparison to those of potyviruses. J Gen Virol 76: 2035–2042
13. Hall BG (2001) Phylogenetic trees made easy. Sinauer, Sunderland, Mass
14. Hall JS, Adams B, Parsons TJ, French R, Lane LC, Jensen SG (1998) Molecular cloning, sequencing, and phylogenetic relationships of a new potyvirus: sugarcane streak mosaic virus, and a reevaluation of the classification of the Potyviridae. Mol Phylogenet Evol 10: 323–332
15. Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22: 160–174
16. Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA 89: 10915–10919
17. Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analyses. Syst Biol 42: 182–192
18. Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogeny. Bioinformatics 17: 754–755
19. Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17: 1244–1245
20. Loytynoja A, Milinkovitch MC (2003) A hidden Markov model for progressive multiple alignment. Bioinformatics 19: 1505–1513
21. Mackenzie AM, Webster DE, Gibbs MJ, Thomas BJ, Gibbs AJ (1999) Australian isolates of ryegrass mosaic rymovirus and their relationships. Arch Virol 144: 309–316

22. Niblett CL, Zagula KR, Calvert LA, Kendall TL, Stark DM, Smith CE, Beachy RN, Lommel SA (1991) cDNA cloning and nucleotide sequence of the wheat streak mosaic virus capsid protein gene. J Gen Virol 72: 499–504
23. Page, RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12: 357–358
24. Rabenstein F, Seifers DL, Schubert J, French R, Stenger DC (2002) Phylogenetic relationships, strain diversity and biogeography of tritimoviruses. J Gen Virol 83: 895–906
25. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European molecular biology open software suite. Trends Genet 16: 276–277
26. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425
27. Salm SN, Rey MEC, Robertson NL, French R, Rabenstein F, Schubert J (1996) Molecular cloning and nucleotide sequencing of the partial genomes of Agropyron and Hordeum mosaic viruses, two members of the Rymovirus genus in the taxonomic family *Potyviridae*. Arch Virol 141: 2115–2127
28. Salm SN, Rey MEC, Rybicki EP (1996) Phylogenetic justification for splitting the Rymovirus genus of the taxonomic family Potyviridae. Arch Virol 141: 2237–2242
29. Sanderson MJ, Shaffer HB (2002) Troubleshooting molecular phylogenetic analyses. Annu Rev Ecol Syst 33: 49–72
30. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502–504
31. Schubert J, Fauquet C, Merits A, Rabenstein F (1999) The complete nucleotide sequence of Ryegrass mosaic potyvirus indicates that it is a recombinant between members of two different genera in the family Potyviridae. J Plant Dis Protect 106: 392–404
32. Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16: 1114–1116
33. Shukla DD, Ward CW, Brunt AA, Berger PH (1998) Potyviridae family. AAB/CMI Descriptions of Plant Viruses, no 366
34. Slykhuis JT, Bell W (1966) Differentiation of Agropyron mosaic, wheat streak mosaic, and a hitherto unrecognized Hordeum mosaic virus in Canada. Can J Bot 44: 1191–1208
35. Stenger DC, French R (2004) Complete nucleotide sequence of *Oat necrotic mottle virus*: a distinct *Tritimovirus* species (family *Potyviridae*) most closely related to *Wheat streak mosaic virus*. Arch Virol 149: 633–640
36. Stenger DC, Hall JS, Choi I-R, French R (1998) Phylogenetic relationships within the family *Potyviridae*: wheat streak mosaic virus and brome streak mosaic virus are not members of the genus *Rymovirus*. Phytopathology 88: 782–787
37. Swofford DL (2003) PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Mass
38. Van Regenmortel MHV, Fauquet CM, Bishop DHL, Carstens EB, Estes MK, Lemon SM, Maniloff J, Mayo MA, McGeoch DJ, Pringle CR, Wickner RB (2000) Virus taxonomy: the classification and nomenclature of viruses. The Seventh Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego
39. Verchot J, Koonin EV, Carrington JC (1991) The 35-kDa protein from the N-terminus of the potyviral polyprotein functions as a third virus-encoded proteinase. Virology 185: 527–535

40. Whelan S, Lio P, Goldman N (2001) Molecular phylogenetics: state-of-the-art methods for looking into the past. Trends Genet 17: 262–272
41. Zagula K, Niblett CL, Robertson NL, French R, Lommel SA (1992) *Potyviridae*: genus *Rymovirus*. Arch Virol Suppl 5: 269–276

Author's address: Roy French, USDA-ARS, 344 Keim Hall, University of Nebraska, Lincoln, NE 68583-0722, U.S.A.; e-mail: rfrench@unlnotes.unl.edu