

# Co-divergence and host-switching in the evolution of tobamoviruses

Anthony H. Stobbe,<sup>1</sup> Ulrich Melcher,<sup>1</sup> Michael W. Palmer,<sup>2</sup>  
Marilyn J. Roossinck<sup>3†</sup> and Guoan Shen<sup>3</sup>

## Correspondence

Ulrich Melcher  
u-melcher-4@alumni.uchicago.  
edu

<sup>1</sup>Department of Biochemistry and Molecular Biology, Oklahoma State University, Stillwater, OK 74078, USA

<sup>2</sup>Department of Botany, Oklahoma State University, Stillwater, OK 74078, USA

<sup>3</sup>Plant Biology Division, The Samuel Roberts Noble Foundation, Ardmore, OK 73401, USA

The proposed phylogenetic structure of the genus *Tobamovirus* supports the idea that these viruses have codiverged with their hosts since radiation of the hosts from a common ancestor. The determinations of genome sequence for two strains of Passion fruit mosaic virus (PafMV), a tobamovirus from plants of the family Passifloraceae (order Malpighiales) from which only one other tobamovirus (*Maracuja mosaic virus*; MarMV) has been characterized, combined with the development of Bayesian analysis methods for phylogenetic inference, provided an opportunity to reassess the co-divergence hypothesis. The sequence of one PafMV strain, PfaMV-TGP, was discovered during a survey of plants of the Tallgrass Prairie Preserve for their virus content. Its nucleotides are only 73% identical to those of MarMV. A conserved ORF not found in other tobamovirus genomes, and encoding a cysteine-rich protein, was found in MarMV and both PafMV strains. Phylogenetic tree construction, using an alignment of the nucleotide sequences of PafMV-TGP and other tobamoviruses resulted in a major clade containing isolates exclusively from rosoid plants. Asterid-derived viruses were exclusively found in a second major clade that also contained an orchid-derived tobamovirus and tobamoviruses infecting plants of the order Brassicales. With a few exceptions, calibrating the virus tree with dates of host divergence at two points resulted in predictions of divergence times of family specific tobamovirus clades that were consistent with the times of divergence of the host plant orders.

Received 17 May 2011

Accepted 1 November 2011

## INTRODUCTION

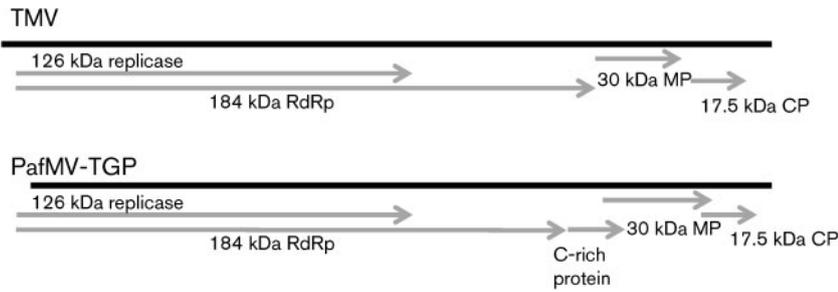
Hallmark virus genes are those whose relatives can be recognized in a wide variety of virus genomes but are not found in any host organism genome (Koonin *et al.*, 2006). Their recognition suggests that viruses have existed since life began. Recent work finding insertions of virus sequences in multiple mammalian genomes (Belyi *et al.*, 2010) supports the antiquity of viruses, and suggests the presence of viruses at least 40–50 million years ago (mya). Consistent with an ancient origin of viruses, the species diversity of angiosperm viruses may have arisen largely during long association of viruses and hosts during the angiosperms' divergence and proliferation. Such diversification is designated 'co-divergence' to emphasize a lesser role of selection in the process. The angiosperm virus genus *Tobamovirus* of the family *Virgaviridae* (Adams *et al.*,

2009) has been proposed to have evolved by co-divergence (Gibbs, 1980, 1999; Lartey *et al.*, 1996).

The type member of the genus *Tobamovirus* (Fauquet *et al.*, 2005), *Tobacco mosaic virus* (TMV) figured prominently in the recognition of the existence of filterable infectious agents now called viruses (Schothof *et al.*, 1999). The TMV genome consists of a positive-sense ssRNA molecule of 6.4 kilonucleotide (knt) that is packaged in a non-enveloped rod-shaped coat approximately 300 nm in length. Particle preparations also contain shorter rods, corresponding to encapsidated subgenomic RNAs that contain the virus origin of assembly (Fukuda *et al.*, 1981). The TMV genome has four ORFs, encoding a 126 kDa replicase component that has methyl-transferase (MT), helicase (Hel) and RNA silencing suppressor domains (Vogler *et al.*, 2007), a 183 kDa protein which is a read-through product of the 126 kDa protein ORF and has an additional RNA-dependent RNA polymerase (RdRp) domain, a movement protein (MP) and a coat protein (CP) (Fig. 1). The 126 kDa and/or the 183 kDa protein and the MP are necessary for TMV intercellular movement (Harries *et al.*, 2009).

†Present address: Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA 16802, USA.

Supplementary material is available with the online version of this paper.



**Fig. 1.** Genome structure of PafMV-TGP compared to that of TMV. Arrows indicate ORFs. Sizes of 120 and 180 kDa polypeptides are approximate due to lack of the 5'-end of the nucleotide sequence.

Co-divergence of tobamovirus lineages with those of their hosts is suggested by apparently slow rates of sequence change and the structure of phylogenetic trees based on both nucleotide and amino acid sequences. Several lines of evidence support a slow rate of accumulation of nucleotide substitutions within species of the tobamoviruses. The initial sequence determination of the TMV genome was performed in two laboratories (Dawson *et al.*, 1986; Goellet *et al.*, 1982) on samples that had been propagated separately in tobacco for 30 years and yet they yielded virtually identical sequences. Sequence comparisons of archival *Tobacco mild green mosaic virus* specimens from wild *Nicotiana glauca* covering a span of close to 100 years showed little evidence of divergence at this timescale (Fraile *et al.*, 1997).

Phylogenetic evidence showing clustering according to host taxonomy also supports the co-divergence hypothesis for tobamovirus species. Initial support was obtained from a comparison of CP amino acid sequences of known tobamoviruses (Gibbs, 1980) and was extended by a comparison of nucleotide sequences of each coding region (Gibbs, 1999; Lartey *et al.*, 1996). Where distinct virus species have been isolated from members of the same plant order (Cucurbitales, Malvales, Fabales, Solanales, Brassicales and Lamiales), almost invariably their sequences have clustered on the same branch of the viral phylogenetic trees (Gibbs, 1999; Lartey *et al.*, 1996; Min *et al.*, 2009). For the members of the orders Cucurbitales, Malvales and Fabales, the branches do not include viruses isolated from other plant orders. Viruses of members of Brassicales and Lamiales cohabit the same branch, designated subgroup III (Lartey *et al.*, 1996) to distinguish it from the branch associated with plants of the Solanales (subgroup I) and other tobamoviruses (subgroup II). A virus isolated from orchids has one genome portion that branches with the viruses from Solanales members and another that is found on the Brassicales–Lamiales branch, suggesting that it is a recombinant (Lartey *et al.*, 1996). Comparative sequence analysis also revealed that all lineages of tobamoviruses except that of the recombinant tobamovirus had impressively uniform rates of sequence evolution, consistent with genetic drift minimally influenced by selective events. Patterns of co-divergence of other virus species with the species of their hosts have been identified (Pérez-Losada *et al.*, 2006). Cases of co-divergence generally suggest nucleotide substitution

rates in the vicinity of  $10^{-8}$  substitutions per site per year (Gibbs *et al.*, 2010).

The availability of nucleotide sequences from numerous dated isolates of single virus species combined with the development of Bayesian phylogenetic analysis has allowed estimation of the rates of nucleotide substitution in a diversity of virus species (Duffy & Holmes, 2008; Fargette *et al.*, 2008; Kang *et al.*, 2009; Moore & Donoghue, 2009; Pérez-Losada *et al.*, 2006; Ramsden *et al.*, 2009). The calculated substitution rates are similar to mutation rates estimated experimentally by analysis of progeny from plants inoculated with cloned sequences. The rates are in the range of  $10^{-3}$  to  $10^{-5}$  substitutions per site per year (Duffy & Holmes, 2008; Fargette *et al.*, 2008). Extrapolation of such rates linearly over time to represent the evolution of species led to the conclusion that current virus taxa arose recently rather than having co-diverged with their hosts. Recently, Bayesian analysis has been extended to selected individual species of the genus *Tobamovirus* and identified a similarly high substitution rate (Pagán *et al.*, 2010) for those species. A similarly high rate of evolution for the virus species in the genus, extrapolated from that analysis, would be inconsistent with the co-divergence hypothesis.

Plants of the family Passifloraceae (order Malpighiales) host a variety of viruses, including the tobamoviruses *Maracuja mosaic virus* (MarMV) (Song *et al.*, 2006) and Passion fruit mosaic virus (PafMV) (Song & Ryu, 2011). An isolate of the latter (PafMV-FL, GenBank accession no. NC\_015552) was originally thought to be a strain of MarMV, MarMV-FL. While this manuscript was in preparation, its sequence was reported (Song & Ryu, 2011), revealing that the virus is a distinct species in the genus *Tobamovirus* and proposing the PafMV name. The Plant Virus Biodiversity and Ecology (PVBE) project focused on determining the distribution of plant viruses in close to 600 native or non-crop plant species in the Tallgrass Prairie Preserve of Osage Co., Oklahoma (TGP) through sequence analysis (Wren *et al.*, 2006). One virus sequence identified in the study suggested that the virus was a strain of PafMV, which we refer to here as PafMV-TGP. Characterization of the PafMV-TGP genome revealed a novel conserved C-rich ORF of unknown function also found in PafMV-FL and MarMV. The availability of these new tobamovirus genomes coupled with those of other

newly sequenced tobamovirus genomes (Min *et al.*, 2006; Srinivasan *et al.*, 2005) allowed a new examination of the co-divergence hypothesis for tobamoviruses.

## RESULTS

### PafMV-TGP

In the PVBE project, viral sequences in plant extracts were enriched either by isolation of dsRNA or by the preparation of virus-like particles (VLP) prior to nucleic acid extraction (producing VLP-VNA; see Methods). Both approaches in identifying viral sequences in plant extracts yielded unequivocal identification of two plants with evidence of the presence of a member of the genus *Tobamovirus*. One plant, 05TGP00580 (sampled on 24 July, 2005 from 36.848° N, 96.420° W), the only sample of the *Passiflora incarnata* species analysed, produced a high percentage of tobamovirus sequence reads in both methods (VLP-VNA, 85.2% of 2931 reads; dsRNA, 66.0% of 280 reads). The other, 07TGP00004 (sampled on 8 June, 2007 from 36.838° N, 96.443° W), was a sample of *Vernonia baldwinii* represented by 0.7% of 597 reads.

Sequences obtained (Supplementary Material, available in JGV Online) were sufficiently abundant to allow the assembly of 6696 nt into one large contig, missing only short sequences (less than 100 nt) at the 5'- and 3'-ends, probably including the first 11 codons of the 5'-most ORF. The high percentages of nucleotide and predicted amino acid identity for PafMV-TGP compared to PafMV-FL (Table 1) clearly identified the two strains as belonging to the same species. Among known tobamovirus genomes, the next closest known relative of the sequence was that of *Maracuja mosaic virus* (MarMV), exhibiting 72.7% nucleotide sequence identity. The PafMV-TGP genome (Fig. 1) contains similar ORFs in a similar order to other members of the genus *Tobamovirus*, but with the following notable differences. A stretch of 374 nt separates the 185 kDa ORF termination codon from the MP initiation codon. The 185K

ORF termination codon is followed, starting four nucleotides 3' of it, by an ORF of 594 nt. It thus overlaps the MP ORF out of frame for 220 nt. This ORF is also present in the PafMV-FL and MarMV genomes although it was not described in the published reports (Song *et al.*, 2006; Song & Ryu, 2011) and is not annotated in GenBank sequences. The ORF can encode a hypothetical cysteine-rich protein located between nucleotide number 4809 and 5403. In a BLASTP search of protein sequences, the amino acid sequence of the predicted product of the ORF lacked similarity to proteins of characterized functions. Conservation of this ORF in both PafMV strains and MarMV suggests that it is a functional ORF. The PafMV-TGP and MarMV MP regions overlap with the CP regions in a different frame for 118 nt, similar to the overlap previously noted for crucifer-associated tobamoviruses (Lartey *et al.*, 1996). The overlap regions involving each of the last three coding regions perhaps account for the higher per cent nucleotide identity for these regions (Table 1). Of 6098 PafMV-TGP positions covered by more than one contig or singleton read, only 29 (0.5%) exhibited evidence of polymorphism. Of those 29, only two were informatively polymorphic (Supplementary Table S1, available in JGV Online). The highest densities of polymorphic positions were in the ORF for the putative C-rich polypeptide and the C-terminal part of the movement protein, 8.4 and 7.5 polymorphisms per knt, respectively, compared to 3.1, 3.2 and 3.7 polymorphisms per knt, respectively, for the MT-Hel domain, the RdRp domain and the CP ORF.

### Phylogenetics

A Bayesian-likelihood tree of the replicase (MT-Hel-RdRp) ORFs of the tobamoviruses (Fig. 2a), clearly defined clades containing the Lamiales-, Solanales-, Fabales-, Malvales- and Cucurbitales-associated lineages. This definition was also obtained by maximum-likelihood analysis (data not shown). The asterid Lamiales-associated clade, corresponding to subgroup III, also included viruses associated with the asterid Ericales and the rosid Brassicales orders. Viruses infecting plants of these orders were

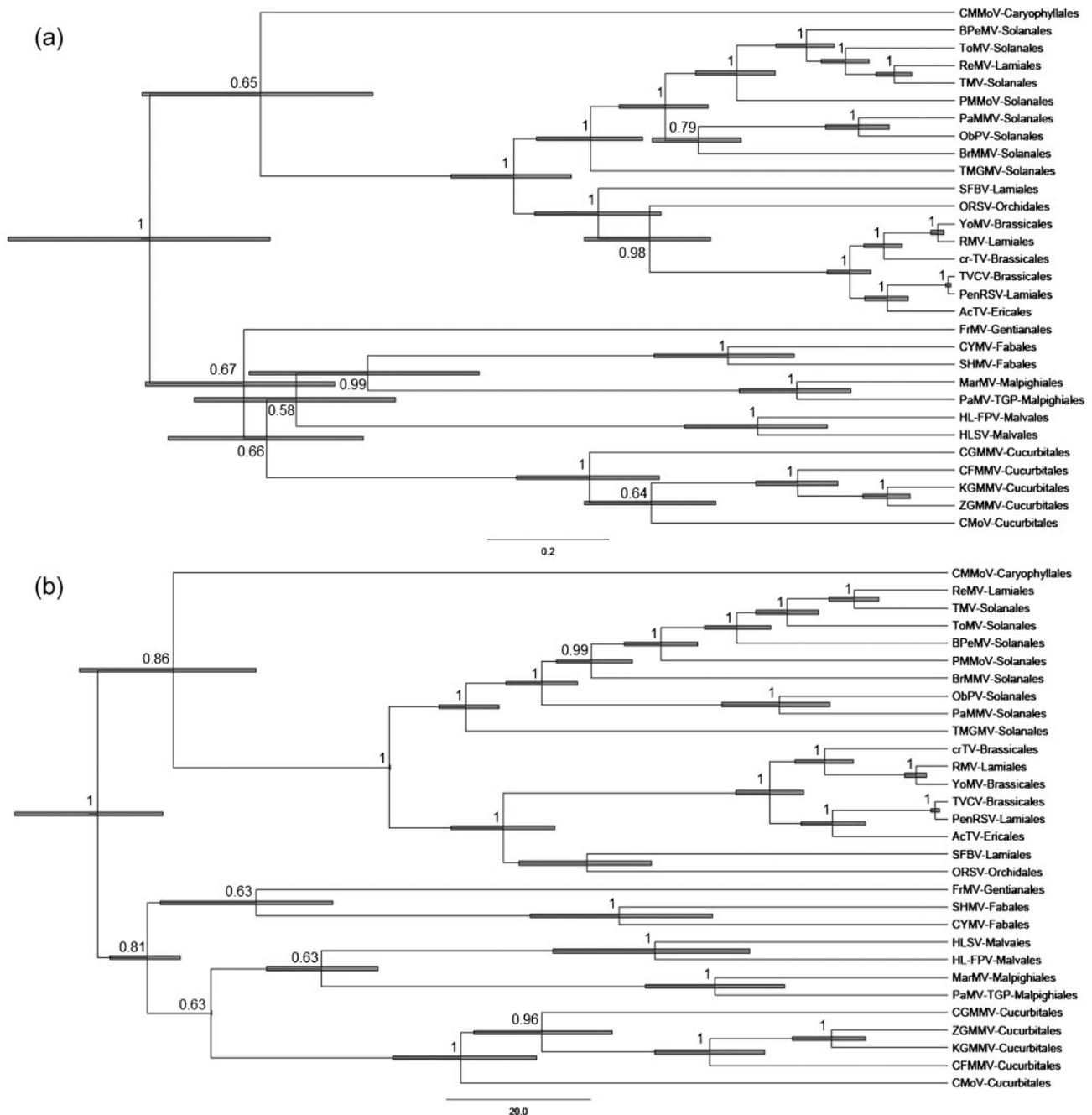
**Table 1.** Comparison of nucleotide and predicted amino acid sequences of two strains of PafMV

NA, Not applicable.

| Region                  | Nucleotide identity* | Amino acid identity* |
|-------------------------|----------------------|----------------------|
| MT-Hel (125K)†          | 94.9 (3062/3267)     | 98.8 (1077/1089)     |
| Replicase (184K)        | 94.1 (4511/4794)     | 98.9 (1583/1601)     |
| Putative C-rich protein | 96.5 (573/594)       | 96.0 (190/198)       |
| MP                      | 96.4 (899/933)       | 99.7 (310/311)       |
| CP                      | 97.8 (522/534)       | 97.8 (174/178)       |
| Overall                 | 94.9                 | NA                   |

\*Percentage identity (no. identical/no. positions) comparing PafMV-TGP (GenBank accession no. JF807914) and PafMV-FL (GenBank accession no. NC\_015552).

†First 89 nt residues of PafMV-TGP missing.



**Fig. 2.** Bayesian-likelihood phylogenetic trees using the replicase gene of tobamoviruses without (a) and with (b) dating priors. The dated tree (b) is in millions of years (bar=20 million years). Posterior probabilities are shown on the nodes. Names of viruses and sources of their sequences are given in Supplementary Table S2.

interspersed in the topology of the subgroup III clade. The Solanales-associated clade, corresponding to subgroup I, included *Rehmannia mosaic virus* (ReMV), a close relative of TMV whose host is a member of the Lamiales. At a deeper level, subgroup I and subgroup III clades clustered together separately from subgroup II viruses. The subgroup I–III branch was subtended by one containing a single member clade consisting of *Cactus mild mottle virus*

(CMMoV, host order Caryophyllales), while the other single member clade, that of *Frangipani mosaic virus* (FrMV, host order Gentianales of the asterid clade) appeared basal to the subgroup II clade. However, the confidence intervals were such that the two single member clades and all order-specific subgroup II branches may have originated at about the same time. Among these branches, PafMV-TGP appeared as sister to MarMV in a

**Table 2.** Analysis of clock-like behaviour of tobamoviral ORFs

NA, Not applicable.

| ORF    | uclid.stdev* (aa) | 95 % HPD      | uclid.stdev (nt) | 95 % HPD    |
|--------|-------------------|---------------|------------------|-------------|
| MT–Hel | 0.331             | 0.235–0.435   | 0.317            | 0.232–0.415 |
| RdRp†  | NA                | NA            | 0.267            | 0.192–0.350 |
| RdRp   | 0.271             | 0.188–0.362   | 0.253            | 0.191–0.326 |
| CP     | 0.228             | 2.96E-3–0.407 | 0.376            | 0.208–0.539 |
| MP     | 0.404             | 0.216–0.612   | 0.477            | 0.297–0.680 |

\*This value is a representation of the heterogeneity of rates found in a tree. ORFs with values close to zero are considered to be clock-like, while those whose values are close to, or above unity are considered not clock-like.

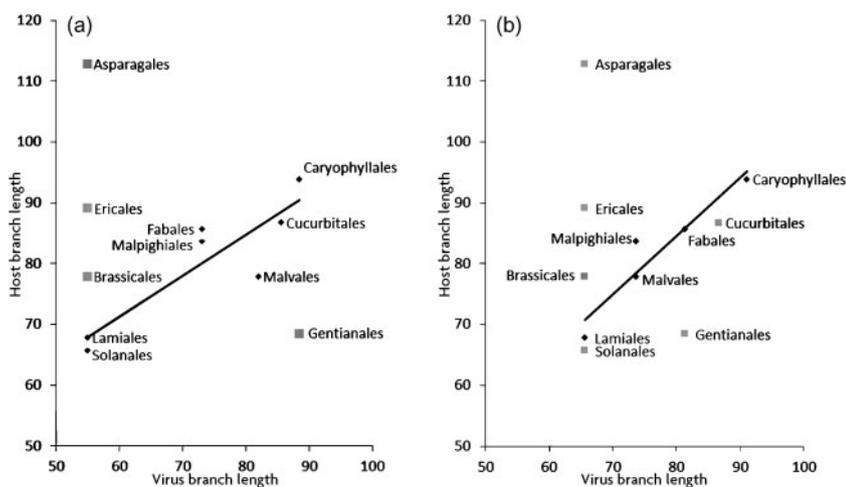
†Uncalibrated tree without priors.

Malpighiales-associated branch. The MT–Hel ORF gave a tree identical in topology and similar in proportions as that generated for the replicase ORF, as expected due to the latter being a continuation of the MT–Hel ORF (Fig. 1). Greater variation was seen in the MP and CP trees, most probably due to the shorter length of their ORFs and greater heterogeneity in their evolutionary rates as shown below.

### Dating

If the substitution rate remains constant throughout the history of the gene, the gene can be said to be clock-like and can be used to extrapolate dates of divergence. The potential clock-like nature of the substitutions within the tobamovirus ORFs was addressed by examining the uclid.stdev parameter of BEAST (SD of the relaxed substitution clock rates; Table 2). This parameter reflects the rate heterogeneity of the lineages. When uclid.stdev > 1, the heterogeneity in rates is high; when uclid.stdev = 0, rates are perfectly clock-like. The uclid.stdev of the nucleotides in the four ORFs ranged from 0.253 to 0.477, with the RdRp ORF being the most clock-like.

If virus clades diverged from each other near the same time that their host orders did, we would expect a linear correlation between age of plant orders and virus clade divergence distances. The ages of seven plant orders (Cucurbitales, Solanales, Malvales, Malpighiales, Lamiales, Caryophyllales and Fabales) were well correlated with the ages of the viral clades associated with them (Fig. 3a,  $R^2=0.795$ , slope=0.670). The strong correlation motivated dating the Bayesian trees, using plant divergence estimates (Magallón & Castillo, 2009). We used a uniform distribution for divergences of plants of the orders Cucurbitales (120.22–120.32 mya) and Solanales (77.42–77.52 mya) from other taxa. Using these points for calibration did not change the topology significantly, but improved the correlation coefficient (Fig. 3b,  $R^2=0.8921$ , slope=0.98). Omitting the Cucurbitales and Solanales from the calculation of correlation did not change the correlation coefficient significantly ( $R^2=0.8994$ , slope=0.96). Estimates of ages of the nodes of the viral replicase derived from this correlation are given in Table 3. The data allowed calculation of the divergence resulting in rosid- and asterid–Caryophyllales-associated groups to have occurred approximately 109–130 mya.



**Fig. 3.** Patristic distances of tobamovirus clades as a function of the divergence times of their associated host plants. Patristic distances were derived from branch lengths from Fig. 2(a, b), respectively, for panels (a) and (b) (without and with dating priors).

**Table 3.** Divergence dates for tobamovirus clades

Deduced by BEAST from the replicase alignment. NA, Not applicable.

| Clade                           | Stem divergence time (mya) | 95 % HPD        | Crown divergence time (mya) | 95 %HPD         |
|---------------------------------|----------------------------|-----------------|-----------------------------|-----------------|
| Asterid–Rosid-associated split  | NA                         | NA              | 118.007                     | 109.018–129.513 |
| Cucurbitales-associated*        | 102.271                    | 102.223–102.318 | 67.584                      | 57.050–77.094   |
| Solanales-associated*           | 77.470                     | 77.424–77.519   | 66.851                      | 62.269–70.655   |
| Malvales-associated             | 86.916                     | 79.043–94.528   | 40.690                      | 27.490–54.818   |
| Lamiales–Brassicales-associated | 77.470                     | 77.424–77.519   | 61.653                      | 54.494–68.918   |
| Caryophyllales-associated       | 107.520                    | 96.068–120.544  | NA                          | NA              |
| Malpighiales-associated         | 86.916                     | 79.043–94.528   | 32.346                      | 22.679–41.924   |
| Fabales-associated              | 96.015                     | 85.340–109.361  | 45.617                      | 32.680–57.961   |
| Gentianales-associated          | 96.015                     | 84.340–109.361  | NA                          | NA              |

\*Clade divergences used as calibration points.

Four virus–host order age relationships were outliers in the correlation plot (Fig. 3): orders Asparaginales, Brassicales, Lamiales and Gentianales. The position of the Asparaginales – *Odontoglossum ringspot virus* (ORSV) correlation can be attributed to the recombinant nature of this virus, mentioned above. Also as mentioned above, the viruses of the members of Brassicales and Ericales are in a clade with viruses derived from the members of Lamiales. FrMV was extrapolated to have separated 16 mya earlier than the plants of the order Gentianales diverged from other asterids.

Where estimatable, crown ages (times since first evidence of divergence in a lineage) for virus branches were an average of 27 million years less than those for the host crown ages (Table 3), except for the order Malvales and their viruses, but were well within the 95% highest posterior density (HPD). Both Malpighiales-associated virus species were isolated from plants of the genus *Passiflora* and would therefore be expected to have diverged later than when the plant orders radiated. The substitution rates of each branch diverging within the interspersed Lamiales–Brassicales-associated clade showed little heterogeneity in their calculated rates of evolution (Supplementary Sequence Alignments zip file, available in JGV Online).

### Bayesian Tip-Significance (BaTS) analysis of association.

If members of the genus *Tobamovirus* did co-diverge alongside their hosts (a form of allopatric speciation), an association of the branching patterns of the primary natural hosts and their viruses should be seen (Kitchen *et al.*, 2011). To test this, the tips of the posterior set of trees found using the above method were labelled with the host order from which the virus was derived. This labelled tree set was tested using BaTS software (Parker *et al.*, 2008), which tests the association of the primary host and the

virus employing three independent statistical tests, association index, parsimony score and maximum monophyletic clade. Both the association index and parsimony score tests showed a strong association of states within the trees. The maximum monophyletic clade test showed strong association ( $P < 0.05$ ) of the orders Cucurbitales-, Malvales-, Fabales- and Malpighiales-associated viruses (Table 4). Viruses of subgroup I (Solanales-associated) were less strongly associated ( $P < 0.5$ ). This is probably due to the inclusion of ReMV, a Lamiales-associated virus. Analysis using the posterior set of trees (PSTs) attained from the CP and MP regions showed subgroup I to have a strong association ( $P < 0.05$ ). Gentianales-, Caryophyllales- and Ericales-derived viruses only had one representative tip and, because of this, were not able to associate with other tips. The Lamiales- and Brassicales-associated viruses did not show any statistical association ( $P = 1$ ).

## DISCUSSION

### Co-divergence

Both RNA and DNA viruses have been argued to have codiverged with their hosts (Pérez-Losada *et al.*, 2006; Wu *et al.*, 2008). The principal observations supporting co-divergence hypotheses in general are the congruence of viral phylogenetic tree reconstructions with those of the host organisms with which the viruses are associated. As the number of characterized tobamoviruses has increased (Min *et al.*, 2006; Song & Ryu, 2011; Srinivasan *et al.*, 2005) the validity of the generalization that virus phylogenetic trees resemble those of their isolation plant hosts has been strengthened. The present analysis supports a co-divergence hypothesis for tobamoviruses strongly. PafMV was confidently assigned to a clade of tobamoviruses associated with hosts in the order Malpighiales (Song & Ryu, 2011; this work). In the survey of plants of the TGP, only a

**Table 4.** Statistical evidence for host association within a tobamovirus clade

Deduced with BaTS using the using the PSTs from the above BEAST analysis.

| Statistic (state)*   | Observed mean | Lower 95 % CI | Upper 95 % CI | Null mean | Lower 95 % CI | Upper 95 % CI | Significance |
|----------------------|---------------|---------------|---------------|-----------|---------------|---------------|--------------|
| AI                   | 1.487         | 1.311         | 1.552         | 2.980     | 2.467         | 3.372         | 0.000        |
| PS                   | 13.000        | 13.000        | 13.000        | 20.426    | 18.968        | 22.000        | 0.000        |
| MC (Ericales)†       | 1.000         | 1.000         | 1.000         | 1.000     | 1.000         | 1.000         | 1.000        |
| MC (Solanales)       | 2.011         | 2.000         | 2.000         | 1.547     | 1.000         | 3.000         | 0.437        |
| MC (Caryophyllales)† | 1.000         | 1.000         | 1.000         | 1.000     | 1.000         | 1.000         | 1.000        |
| MC (Fabales)         | 2.000         | 2.000         | 2.000         | 1.015     | 1.000         | 1.000         | 0.015        |
| MC (Brassicales)     | 1.000         | 1.000         | 1.000         | 1.058     | 1.000         | 1.782         | 1.000        |
| MC (Cucurbitales)    | 5.000         | 5.000         | 5.000         | 1.221     | 1.000         | 2.000         | 0.001        |
| MC (Gentianales)†    | 1.000         | 1.000         | 1.000         | 1.000     | 1.000         | 1.000         | 1.000        |
| MC (Malvales)        | 2.000         | 2.000         | 2.000         | 1.018     | 1.000         | 1.000         | 0.014        |
| MC (Malpighiales)    | 2.000         | 2.000         | 2.000         | 1.020     | 1.000         | 1.000         | 0.019        |
| MC (Orchidales)†     | 1.000         | 1.000         | 1.000         | 1.000     | 1.000         | 1.000         | 1.000        |
| MC (Lamiales)        | 1.000         | 1.000         | 1.000         | 1.117     | 1.000         | 2.000         | 1.000        |

\*Association index (AI), parsimony score (PS) and maximum monophyletic clade (MC).

†Only one state included.

specimen of the genus *Passiflora* accumulated PafMV-TGP to high levels. Plants of an order are most often hosts to a single monophyletic clade of tobamoviruses found in that order.

The virus tree (Fig. 2) mirrored the plant species tree in most regards. Malvales-, Cucurbitales-, Fabales- and Malpighiales-associated viruses each formed monophyletic clades. All four of their plant orders are rosids and no asterid-associated viruses were in the branch that included these clades, indicating that also supra-order association exists. Two single member clades, CMMoV and FrMV, appeared to diverge from common ancestors with the rosid clade and a clade containing all other known asterid-associated viruses. The phylogenetic distances of the points of divergence of seven of nine order-specific virus clades were significantly correlated with the ages of the plant orders (Fig. 3b). The oldest branch point within each order-specific virus clade occurred within 10 million years after diversification of the order (Table 3). The splitting of virus lineages following splitting of the host lineage is consistent with co-divergence. Furthermore, and most importantly, BaTS analysis showed (Table 4) that the virus tree topology and its association with host taxonomy is highly unlikely to have resulted from random processes. Combined, these observations suggest strongly that the apparent relationship of virus clades with the phylogeny of their natural hosts reflects an important evolutionary phenomenon.

### Alternatives to co-divergence

An evolutionary relationship underlying tree similarity requires three assumptions. It assumes that a virus species' host of initial isolation reflects the hosts in which it spent most of its time evolving. Consistent with the assumption,

annotations of known natural hosts of tobamoviruses in the International Committee for the Taxonomy of Viruses (2006) database reveal no taxonomically widespread distribution of natural hosts. In the TGP plant community, PafMV-TGP was found only twice. Of about 450 specimens of six frequently sampled species, only one was definitively positive for the virus. Of 550 plant species tested, only two yielded evidence of the virus. The *P. incarnata* specimen stood out as having a high concentration of the tobamovirus without any obvious symptoms of infection, suggesting that it was the source of the virus found in the other plant. *P. incarnata* was also the probable source of PafMV-FL cuttings, which may have been transported to Florida from Arkansas (St Hill *et al.*, 1992), the state neighbouring that of the PVBE study. These observations suggest that PafMV established a long-term productive association with one host lineage, while still occurring in plants of other lineages in ways that do not contribute to the evolution of the virus. TGP viruses such as the *Asclepias* asymptomatic tymovirus (Min *et al.*, 2011) and *Asclepias* virus TGP-2 (V. Thapa and others, unpublished data), a proposed member of the Secoviridae, were detected at substantial levels in many host species, although highest populations were in *Asclepias viridis*. Thus, assuming that the host of initial isolation is an indicator of the host plant lineage in which tobamoviruses evolved is reasonable, but may not be valid for other viruses.

A second assumption in asserting co-divergence based on tree similarity is that co-divergence could be accompanied by occasional plant lineage jumps. Jumps may have happened several times in tobamovirus evolution. Yet, some apparent jumps in Fig. 2 are not statistically supportable. The order of divergence of order-specific virus clades associated with the orders Fabales, Cucurbitales, Malpighiales and Malvales is

uncertain as witnessed by the discrepancies between the dated and undated trees (Fig. 2a, b). One incontrovertible exception to absolute tree congruence is ReMV, a virus of a plant species in the Lamiales whose closest relative is the Solanales-associated TMV. A second exception is FrMV whose isolation host was from an asterid order, yet the virus branched early from the rosoid lineage. Most interestingly, in the asterid-associated branch, all Solanales-associated viruses included in the study occur on one sub-branch (subgroup I), also occupied by ReMV, but the other sub-branch (subgroup III) contains viruses from multiple orders: the remaining Lamiales, the rosoid Brassicales and ORSV (discussed later). Close relatives of these viruses have also been identified from the asterid orders Solanales [*Petunia* (Sabanadzovic *et al.*, 2008)] and Ericales [(*Actinidia* and *Impatiens*) (Chavan *et al.*, 2009; Heinze *et al.*, 2006)]. Within this mixed-order sub-branch no substructure according to plant host occurs, suggesting that this lineage acquired the ability to be successful in multiple hosts. Placement of this sub-branch as sister to the primarily asterid Solanales-associated sub-branch suggests that the lineage arose in the astrid Lamiales and subsequently gained the ability to survive and spread in other lineages. That tobamoviruses from Brassicales and Lamiales do not form separate clades suggests that this new ability was attained after the divergence of subgroup III from subgroup I.

A third assumption in asserting co-divergence is that rates of nucleotide substitution are much slower in the process of evolution of viral species than they are during evolution within a species. In this study, as has been done in others (Rector *et al.*, 2007), calibrating the viral tree using estimated dates of plant order divergence (Magallón & Castillo, 2009) gave substitution rates of the order of  $10^{-8}$  to  $10^{-9}$  substitutions per site per year a rate compatible with the observation that sequences of tobamoviruses from century-old herbarium specimens are not appreciably different from modern sequences (Fraile *et al.*, 1997). Slow evolutionary substitution rates for species evolution have also been proposed for other viruses (Gibbs *et al.*, 2010; Wu *et al.*, 2008). Rates estimated by Bayesian analysis for the evolution of sequences within species using dated isolates are orders of magnitude higher than the interspecies inferences assuming co-divergence (Harkins *et al.*, 2009; Pagán *et al.*, 2010). For tobamoviruses, Pagán *et al.* (2010) found rates of  $10^{-4}$  to  $10^{-5}$  substitutions per site per year by comparing tobamovirus samples of individual species from the asterid clade and cucumber green mottle mosaic virus from the past 60 years. Using just CP sequences of selected species, they extrapolated a maximum predicted age of the last common ancestor of known tobamoviruses of  $10^5$  years rather than the  $10^8$  years inferred from the co-divergence hypothesis. In support of the younger estimate, rates of the magnitude  $10^{-4}$  to  $10^{-5}$  substitutions per site per year have also been obtained upon inoculation of plants with cloned genomes followed by later harvest and sequencing (Ge *et al.*, 2007; Schneider & Roossinck, 2001). The Pagán *et al.* (2010) study of genus evolution differed

from the one presented here in that it focused on the smallest section of the tobamovirus genome (less than 10 % of the total length), rather than the replicase and included only one species outside the Solanales–Lamiales associated clade as defined in the present study.

For three reasons, it is risky to extrapolate results from the Bayesian analysis of substitution during intraspecies evolution linearly to the events that gave rise to those species. First, substitution profiles for within species variation differ from those of between species variation. The types of substitutions dominating substitution profiles of tobamoviruses vary with the taxonomic level of the comparison (Melcher, 2010). G↔A and T↔C transitions predominate among recently diverged pairs of sequences but accumulate at rates similar to those of other substitutions with more diverged pairs. Second, observations suggest that the nucleotide sequences of viruses in their natural hosts are stable over large evolutionary scales (Acosta-Leal *et al.*, 2011). Purifying selection may be particularly strong and bottlenecks may be comparatively wide. These views are consistent with the comparatively low density of polymorphisms among the sequences retrieved for PafMV-TGP. Third, the apparent discrepancy between interspecies and intraspecies substitution rates can be reconciled easily by recognizing that the best evolution models used in Bayesian estimations are usually those that invoke a category of sites that are invariant. However, the proportion of sites that are actually invariant will decrease with increasing phylogenetic distance. Substitution rates within the evolution of individual species focus on a limited number of changing sites. These play only a small role in phylogenetic inference at the interspecies level because the sites are close to being saturated with changes at the longer times. The short-term invariant sites do undergo substitution during interspecies evolution and it is those substitutions that are important in the phylogenetic reconstruction of species evolution. An additional manner by which studies of historical strains can lead to misleading rates is that strains of a virus may disappear for a time and then reappear in a later year, such as has been reported for TGP carmovirus 3 (Scheets *et al.*, 2011), such that the differences between them do not reflect sequence evolution on the timescale postulated but on a longer one resulting from evolution of strains. It bears noting, correspondingly, that the wide difference in rates among sites means that Bayesian estimates of divergence times based on co-divergence with hosts are exaggeratedly high for the most recent nodes.

A fourth assumption of the tobamovirus–host co-divergence hypothesis is that an ancestral tobamovirus existed before the dicotyledonous plants radiated. The recent analysis of the genome of *Chara australis* virus (CAV), a virus associated with a brown alga (Gibbs *et al.*, 2011) suggests this assumption is warranted. The CP gene of CAV is homologous to those of known tobamoviruses, but is about 1.7-fold more ancient than the radiation of tobamoviruses of embryophyta radiating. Parsimony analysis of

the CP ORF nucleotide sequence (data not shown), but not the amino acid sequences (Gibbs *et al.*, 2011), placed the CAV branch in the vicinity of the origin of tobamovirus diversification. This origin can be deduced from the tree (Fig. 2b) from the region where the confidence limits of many of the order-specific branches overlap. The Fabales-, Malpighiales- Cucurbitales-, Caryophyllales- and Gentianales-associated virus clades and the combined Solanales-Lamiales-associated virus clade have similarly deep branches, suggesting that the root of the dicotyledonous tobamovirus tree is in this region. However, it is important to recognize that there are many plant orders from which no tobamovirus has been detected to date. Since several orders harbouring tobamoviruses have only recently been reported, this absence could be due to inadequate sampling. An alternative, suggested earlier for the order Brassicales (Lartey *et al.*, 1996), is that the original tobamovirus lineages in those orders now without tobamoviruses have died out.

### Generation of co-divergence

If reasonable assumptions support the designation of the similarity between viral and natural host trees as co-divergence, one must ask: what is the cause of co-divergence? It is not the result of limited sampling. Early suggestions of co-divergence only identified two rosid-associated viruses, one in the order Fabales and the other in the order Cucurbitales. Since then four additional Cucurbitales-associated viruses and one Fabales infecting virus have been described and new branches for Malvales-associated viruses and Malpighiales-associated virus each have more than one clade member. While this manuscript was being prepared, a report of a tobamovirus from a plant of the order Caryophyllales appeared with phylogenetic placement as sister to CMMoV, also from a Caryophyllales plant (Kim *et al.*, 2011). Additional Solanales-derived and Lamiales-associated viruses have, for the most part, been placed on the appropriate branches.

A second argument against attributing tree similarity to co-divergence is that the apparent coincidences of host and viral phylogenetic trees reflect adaptation of the virus to a specific host environment (Holmes, 2008). Viruses in similar host environments (same order, for example) would therefore naturally be more similar than viruses from diverse environments (different orders, for example). While it has been shown that adaptive sequence changes occur when a virus is experimentally transferred from one host to another, the number of such changes is usually small and thus of minor influence in shaping phylogenetic trees (Wallis *et al.*, 2007). Tobamoviruses provide at least two examples of the failure of adaptation to provide major phylogenetic signals. The orchid-associated tobamovirus, ORSV, has been shown to be a recombinant and some acceleration of evolution associated with recombination was detected (Lartey *et al.*, 1996). Nevertheless, adequate phylogenetic signal remains in the recombined segments

to place them with high confidence in known tobamovirus clades. Second, viruses of the Lamiales-Brassicales-associated virus clade infect both rosids and asterids. There are two possible explanations. One requires at least two distinct host 'jumps' from hosts in the Lamiales to hosts in the Brassicales. However, such jumps, if they occurred, could not have been accompanied by any adaptive changes since sequences of tobamoviruses isolated from plants of the order Lamiales are virtually identical to those isolated from plants of the order Brassicales. The alternative explanation is that an earlier event allowed the subgroup to infect plants of multiple orders. The evolutionary rate of a virus is expected to increase after a host jump (Smith *et al.*, 2009) because adaptive substitutions are selected for after the jump. Consistent with this view, jumps in host species are known to increase the intraplant diversity of a virus population (Schneider & Roossinck, 2001). However, the evolutionary rates for each branch within subgroup III do not suggest any specific host crossing event, but instead that the subgroup gained the ability to infect both rosids and asterids approximately between 61 and 21 mya. Because adaptation is unlikely to have played a major role in causing co-divergence, we suggest the following model. Before the dicotyledonous orders radiated, there were large populations of their last common ancestor. These ancestors harboured large populations of diverse copies of a tobamovirus, the diversity being derived by gradual genetic drift from a common ancestor. The large population and the environmental conditions at the time increased the probability of multiple near simultaneous mutations occurring in the same genome, allowing it to establish a new lineage.

## METHODS

**PVBE methods.** Plant sampling and the isolation and extraction of VLP-VNA have been described previously (Melcher *et al.*, 2008), as has the extraction of dsRNA from these plant samples, its conversion to dsDNA and its subsequent sequencing (Roossinck *et al.*, 2010). VLP-VNA sequences reported here were determined as described for dsRNA (Roossinck *et al.*, 2010). Assembled sequences of the putative PafMV-TGP inferred from the contig data were used to identify additional sequences of these viruses among unassembled sequences. These singleton sequences were used in building the consensus sequence. Because of the chance that an occasional read could have been assigned mistakenly to the wrong plant sample in a sequencing pool, a plant sample was called positive using the following criteria: the virus reads were >0.4% of the total reads of the sample, and there was confidence that the read tags were not misread (leading to attribution of the read to the wrong sample).

**Phylogenetic methods.** Sequences for this study were retrieved from GenBank (see Supplementary Table S2, available in JGV Online). CLUSTAL W was used to create a multiple sequence alignment using default values (Larkin *et al.*, 2007). Alignments were viewed using BioEdit (Hall, 1999). Initial studies to place PafMV-TGP into a subgroup used the PHYLIP package. Parsimonious and distance trees were made with *Tobacco rattle virus* (TRV) as an out-group. The evolutionary rates of the tobamoviruses were found by using Bayesian methods with BEAST version 1.5.3 and a most likely tree was made (Drummond & Rambaut, 2007). Log files were viewed to ascertain

convergence using Tracer v1.5. Trees were viewed using FigTree v1.3.0. XML files for BEAST are available in the Supplementary Sequence Alignments zip file. All priors for the undated tree were determined by BEAST. For the dated trees, calibration dates were chosen using the speciation times for the order Cucurbitales and Solanales hosts (120.22–120.32 and 77.42–77.52 mya, respectively) (Magallón & Castillo, 2009) and used a uniform prior distribution. A GTR model of evolution with a gamma distribution of rates and invariant sites was found to be the most probable model via JModelTest (Posada, 2008) and was used in this study. All other priors were determined by BEAST. Four ORFs were analysed with BEAST, the 120 kDa ORF that encodes the MT–Hel protein, the 180 kDa replicase ORF, which includes the MT–Hel and RdRp, the MP ORF and the CP ORF. Each of these was run to 10 million states with a burn-in of 1 million. PSTs were manually edited to include the host-derived states (see Table 4 for states) and a burn-in of 1 million states. These PSTs were then analysed using BaTS with 1000 replicates.

## ACKNOWLEDGEMENTS

The work was supported by funding from NSF-EPSCoR (EPS-0447262), USDA-NIFA (2010-85605-20542) and the Oklahoma Agricultural Experiment Station whose Director has approved the manuscript for publication. The authors thank Plant Virus Biodiversity and Ecology colleagues for the gathering of materials and assistance in nucleic acid preparation, Graham B. Wiley and Bruce A. Roe of the Advanced Center for Genome Technology for 454 sequencing, Andrew Doust, Adrian Gibbs and Akhtar Ali for reading drafts of the manuscript, and Mike Deom and Adrian Gibbs for providing sequences.

## REFERENCES

- Acosta-Leal, R., Duffy, S., Xiong, Z., Hammond, R. W. & Elena, S. F. (2011). Advances in plant virus evolution: translating evolutionary insights into better disease management. *Phytopathology* **101**, 1136–1148.
- Adams, M. J., Antoniw, J. F. & Kreuze, J. (2009). *Virgaviridae*: a new family of rod-shaped plant viruses. *Arch Virol* **154**, 1967–1972.
- Belyi, V. A., Levine, A. J. & Skalka, A. M. (2010). Sequences from ancestral single stranded DNA viruses in vertebrate genomes: the *Parvoviridae* and *Circoviridae* are more than 40–50 million years old. *J Virol* **84**, 12458–12462.
- Chavan, R. R., Pearson, M. N. & Cohen, D. (2009). Partial characterisation of a novel *Tobamovirus* infecting *Actinidia chinensis* and *A. deliciosa* (Actinidiaceae) from China. *Eur J Plant Pathol* **124**, 247–259.
- Dawson, W. O., Beck, D. L., Knorr, D. A. & Grantham, G. L. (1986). cDNA cloning of the complete genome of tobacco mosaic virus and production of infectious transcripts. *Proc Natl Acad Sci U S A* **83**, 1832–1836.
- Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214.
- Duffy, S. & Holmes, E. C. (2008). Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus *Tomato yellow leaf curl virus*. *J Virol* **82**, 957–965.
- Fargette, D., Pinel, A., Rakotomalala, M., Sangu, E., Traoré, O., Sérémé, D., Sorho, F., Issaka, S., Hébrard, E. & other authors (2008). *Rice yellow mottle virus*, an RNA plant virus, evolves as rapidly as most RNA animal viruses. *J Virol* **82**, 3584–3589.
- Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U. & Ball, L. A. (editors) (2005). *Virus Taxonomy*. London: Elsevier.
- Fraile, A., Escriu, F., Aranda, M. A., Malpica, J. M., Gibbs, A. J. & García-Arenal, F. (1997). A century of tobamovirus evolution in an Australian population of *Nicotiana glauca*. *J Virol* **71**, 8316–8320.
- Fukuda, M., Meshi, T., Okada, Y., Otsuki, Y. & Takebe, I. (1981). Correlation between particle multiplicity and location on virion RNA of the assembly initiation site for viruses of the tobacco mosaic virus group. *Proc Natl Acad Sci U S A* **78**, 4231–4235.
- Ge, L., Zhang, J., Zhou, X. & Li, H. (2007). Genetic structure and population variability of tomato yellow leaf curl China virus. *J Virol* **81**, 5902–5907.
- Gibbs, A. (1980). How ancient are the tobamoviruses? *Intervirology* **14**, 101–108.
- Gibbs, A. (1999). Evolution and origins of tobamoviruses. *Philos Trans R Soc Lond B Biol Sci* **354**, 593–602.
- Gibbs, A. J., Fargette, D., García-Arenal, F. & Gibbs, M. J. (2010). Time – the emerging dimension of plant virus studies. *J Gen Virol* **91**, 13–22.
- Gibbs, A. J., Torronen, M., Mackenzie, A. M., Wood, J. T., II, Armstrong, J. S., Kondo, H., Tamada, T. & Keese, P. L. (2011). The enigmatic genome of *Chara australis* virus. *J Gen Virol* **92**, 2679–2690.
- Goelet, P., Lomonosoff, G. P., Butler, P. J., Akam, M. E., Gait, M. J. & Karn, J. (1982). Nucleotide sequence of tobacco mosaic virus RNA. *Proc Natl Acad Sci U S A* **79**, 5818–5822.
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**, 95–98.
- Harkins, G. W., Delpont, W., Duffy, S., Wood, N., Monjane, A. L., Owor, B. E., Donaldson, L., Sauntally, S., Triton, G. & other authors (2009). Experimental evidence indicating that mastreviruses probably did not co-diverge with their hosts. *Virol J* **6**, 104.
- Harries, P. A., Park, J.-W., Sasaki, N., Ballard, K. D., Maule, A. J. & Nelson, R. S. (2009). Differing requirements for actin and myosin by plant viruses for sustained intercellular movement. *Proc Natl Acad Sci U S A* **106**, 17594–17599.
- Heinze, C., Lesemann, D. E., Ilmberger, N., Willingmann, P. & Adam, G. (2006). The phylogenetic structure of the cluster of tobamovirus species serologically related to ribgrass mosaic virus (RMV) and the sequence of streptocarpus flower break virus (SFBV). *Arch Virol* **151**, 763–774.
- Holmes, E. C. (2008). Evolutionary history and phylogeography of human viruses. *Annu Rev Microbiol* **62**, 307–328.
- International Committee for the Taxonomy of Viruses (2006) ICTVdB - *The Universal Virus Database*, version 4. [http://ictvdb.bio-mirror.cn/ictv/fs\\_tobam.htm#Genus1/](http://ictvdb.bio-mirror.cn/ictv/fs_tobam.htm#Genus1/)
- Kang, H. J., Bennett, S. N., Sumibcay, L., Arai, S., Hope, A. G., Mocz, G., Song, J. W., Cook, J. A. & Yanagihara, R. (2009). Evolutionary insights from a genetically divergent hantavirus harbored by the European common mole (*Talpa europaea*). *PLoS ONE* **4**, e6149.
- Kim, N. R., Hong, J. S., Song, Y. S., Chung, B. N., Park, J. W. & Ryu, K. H. (2011). The complete genome sequence of a member of a new species of tobamovirus (rattail cactus necrosis-associated virus) isolated from *Aporcactus flagelliformis*. *Arch Virol* (In press)
- Kitchen, A., Shackleton, L. A. & Holmes, E. C. (2011). Family level phylogenies reveal modes of macroevolution in RNA viruses. *Proc Natl Acad Sci U S A* **108**, 238–243.
- Koonin, E. V., Senkevich, T. G. & Dolja, V. V. (2006). The ancient virus world and evolution of cells. *Biol Direct* **1**, 29.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. & other authors (2007). CLUSTAL W and CLUSTAL\_X version 2.0. *Bioinformatics* **23**, 2947–2948.

- Lartey, R. T., Voss, T. C. & Melcher, U. (1996). Tobamovirus evolution: gene overlaps, recombination, and taxonomic implications. *Mol Biol Evol* **13**, 1327–1338.
- Magallón, S. & Castillo, A. (2009). Angiosperm diversification through time. *Am J Bot* **96**, 349–365.
- Melcher, U. (2010). Assessing constancy of substitution rates in viruses over evolutionary time. *BMC Bioinformatics* **11** (Suppl 6), S3.
- Melcher, U., Muthukumar, V., Wiley, G. B., Min, B. E., Palmer, M. W., Verchot-Lubicz, J., Ali, A., Nelson, R. S., Roe, B. A. & other authors (2008). Evidence for novel viruses by analysis of nucleic acids in virus-like particle fractions from *Ambrosia psilostachya*. *J Virol Methods* **152**, 49–55.
- Min, B. E., Chung, B. N., Kim, M. J., Ha, J. H., Lee, B. Y. & Ryu, K. H. (2006). Cactus mild mottle virus is a new cactus-infecting tobamovirus. *Arch Virol* **151**, 13–21.
- Min, B. E., Song, Y. S. & Ryu, K. H. (2009). Complete sequence and genome structure of cactus mild mottle virus. *Arch Virol* **154**, 1371–1374.
- Min, B.-E., Feldman, T. S., Ali, A., Wiley, G., Muthukumar, V., Roe, B. A., Roossinck, M., Melcher, U., Palmer, M. W. & Nelson, R. S. (2012). Molecular characterization, ecology and epidemiology of a novel tymovirus in *Asclepias viridis* from Oklahoma. *Phytopathology* (In press)
- Moore, B. R. & Donoghue, M. J. (2009). A Bayesian approach for evaluating the impact of historical events on rates of diversification. *Proc Natl Acad Sci U S A* **106**, 4307–4312.
- Pagán, I., Firth, C. & Holmes, E. C. (2010). Phylogenetic analysis reveals rapid evolutionary dynamics in the plant RNA virus genus *Tobamovirus*. *J Mol Evol* **71**, 298–307.
- Parker, J., Rambaut, A. & Pybus, O. G. (2008). Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect Genet Evol* **8**, 239–246.
- Pérez-Losada, M., Christensen, R. G., McClellan, D. A., Adams, B. J., Viscidi, R. P., Demma, J. C. & Crandall, K. A. (2006). Comparing phylogenetic codivergence between polyomaviruses and their hosts. *J Virol* **80**, 5663–5669.
- Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol Biol Evol* **25**, 1253–1256.
- Ramsden, C., Holmes, E. C. & Charleston, M. A. (2009). Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for co-divergence. *Mol Biol Evol* **26**, 143–153.
- Rector, A., Lemey, P., Tachezy, R., Mostmans, S., Ghim, S. J., Van Doorslaer, K., Roelke, M., Bush, M., Montali, R. J. & other authors (2007). Ancient papillomavirus-host co-speciation in *Felidae*. *Genome Biol* **8**, R57.
- Roossinck, M. J., Saha, P., Wiley, G. B., Quan, J., White, J. D., Lai, H., Chavarria, F., Shen, G. A. & Roe, B. A. (2010). Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. *Mol Ecol* **19** (Suppl 1), 81–88.
- Sabanadzovic, S., Abou Ghanem-Sabanadzovic, N., Henn, A. & Lawrence, A. (2008). Characterization of a petunia strain of Turnip vein-clearing virus. *J Plant Pathol* **90**, 505–509.
- Scheets, K., Blinkova, O., Melcher, U., Palmer, M. W., Wiley, G. B., Ding, T. & Roe, B. A. (2011). Detection of members of the *Tombusviridae* in the Tallgrass Prairie Preserve, Osage County, Oklahoma, USA. *Virus Res* **160**, 256–263.
- Schneider, W. L. & Roossinck, M. J. (2001). Genetic diversity in RNA virus quasispecies is controlled by host-virus interactions. *J Virol* **75**, 6566–6571.
- Schothof, K.-B. G., Shaw, J. G. & Zaitlin, M. (1999). *Tobacco Mosaic Virus, One Hundred Years of Contributions to Virology*. St. Paul: APS Press.
- Smith, G. J. D., Vijaykrishna, D., Bahl, J., Lycett, S. J., Worobey, M., Pybus, O. G., Ma, S. K., Cheung, C. L., Raghvani, J. & other authors (2009). Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* **459**, 1122–1125.
- Song, Y. S. & Ryu, K. H. (2011). The complete genome sequence and genome structure of passion fruit mosaic virus. *Arch Virol* **156**, 1093–1095.
- Song, Y. S., Min, B. E., Hong, J. S., Rhie, M. J., Kim, M. J. & Ryu, K. H. (2006). Molecular evidence supporting the confirmation of Maracuja mosaic virus as a species of the genus *Tobamovirus* and production of an infectious cDNA transcript. *Arch Virol* **151**, 2337–2348.
- Srinivasan, K. G., Min, B. E., Ryu, K. H., Adkins, S. & Wong, S. M. (2005). Determination of complete nucleotide sequence of *Hibiscus latent Singapore virus*: evidence for the presence of an internal poly(A) tract. *Arch Virol* **150**, 153–166.
- St Hill, A. A., Zettler, F. W., Elliott, M. S., Persent, M. A., Li, R. H. & Bird, J. (1992). Presence of passiflora latent virus and a serologically distinct strain of Maracuja mosaic virus in *Passiflora* spp. in Florida. *Plant Dis* **76**, 843–847.
- Vogler, H., Akbergenov, R., Shivaprasad, P. V., Dang, V., Fasler, M., Kwon, M. O., Zhanybekova, S., Hohn, T. & Heinlein, M. (2007). Modification of small RNAs associated with suppression of RNA silencing by tobamovirus replicase protein. *J Virol* **81**, 10379–10388.
- Wallis, C. M., Stone, A. L., Sherman, D. J., Damsteegt, V. D., Gildow, F. E. & Schneider, W. L. (2007). Adaptation of plum pox virus to a herbaceous host (*Pisum sativum*) following serial passages. *J Gen Virol* **88**, 2839–2845.
- Wren, J. D., Roossinck, M. J., Nelson, R. S., Scheets, K., Palmer, M. W. & Melcher, U. (2006). Plant virus biodiversity and ecology. *PLoS Biol* **4**, e80.
- Wu, B. L., Melcher, U., Guo, X. Y., Wang, X. F., Fan, L. J. & Zhou, G. H. (2008). Assessment of codivergence of mastreviruses with their plant hosts. *BMC Evol Biol* **8**, 335.