98% identical, 100% wrong: per cent nucleotide identity can lead plant virus epidemiology astray

Siobain Duffy* and Yee Mey Seah

Department of Ecology, Evolution and Natural Resources, School of Environmental and Biological Sciences, Rutgers, The State University of New Jersey, New Brunswick, NJ 08901, USA

Short-form publications such as *Plant Disease* reports serve essential functions: the rapid dissemination of information on the geography of established plant pathogens, incidence and symptomology of pathogens in new hosts, and the discovery of novel pathogens. Many of these sentinel publications include viral sequence data, but most use that information only to confirm the virus' species. When researchers use the standard technique of per cent nucleotide identity to determine that the new sequence is closely related to another sequence, potentially erroneous conclusions can be drawn from the results. Multiple introductions of the same pathogen into a country are being ignored because researchers know fast-evolving plant viruses can accumulate substantial sequence divergence over time, even from a single introduction. An increased use of phylogenetic methods in short-form publications could speed our understanding of these cryptic second introductions and aid in control of epidemics.

Keywords: BLAST; phylogeny; biogeography; molecular epidemiology; per cent nucleotide identity

1. INTRODUCTION

In the mid-1990s, the emerging and damaging tomato yellow leaf curl virus (TYLCV) was found in tomato plants in the Caribbean. This Old World virus had never before been seen in the New World and quickly spread to other Caribbean islands, to eastern Mexico and to Florida. From Florida it spread North and West, reaching Alabama in 2005 (Akad *et al.* 2007) and Texas in 2006 (Isakeit *et al.* 2007). At the same time, tomato plants in western Mexico were succumbing to TYLCV (Brown & Idris 2006). As the western Mexican sequences were 98 per cent identical to those from the eastern Caribbean, the infections in western Mexico were thought to be an extension of the initial introduction (Idris *et al.* 2007). Despite this substantial per cent nucleotide identity (PNI), subsequent phylogenetic analysis revealed that these western Mexican sequences were more closely related (greater than 99%) to TYLCV in Asia than in other North American isolates and represented a second introduction of this exotic virus into North America (Duffy & Holmes 2007).

If plant viruses were typically introduced into new locations once and only once, then the different, more complete perspective that phylogenetics provided for TYLCV infections in the New World would be interesting, but of ultimately limited value. We know, however, that multiple introductions are a

frequent occurrence. In the case of TYLCV in the New World, a third introduction, of a mild TYLCV isolate into Venezuela, has also been documented (Zambrano *et al.* 2007).

The possibility of overlooking a second or third introduction of the virus into a country or area stems from the temptation to compare the newly obtained sequence to those of viruses previously sequenced from the same country or from nearby countries. However, infected plant material and disease vectors are inadvertently traded around the world and phylogenies of genes or genomes of individual species of plant viruses often reveal that viruses from distant geographical areas are closely related to one another. For example, outbreaks of iris yellow spot virus in onions from the American state of Georgia are very closely related to strains circulating in Peru (Nischwitz *et al.* 2007). Some pathogens move frequently and are repeatedly re-introduced to certain geographical areas. For instance, cassava mosaic begomoviruses have migrated from eastern Africa to western Africa at least twice (Ndunguru *et al.* 2005), while the maize-adapted maize streak virus A has frequently moved around Africa (Varsani *et al.* 2009). Although incomplete sampling of the diversity of plant viruses hampers definitive source tracking (Moury *et al.* 2006), attempting to find the origin of novel viral sequences can become a useful standard in the field. Placing novel viral sequences into their appropriate phylogeographical context can identify infection source countries, help trace back how the pathogens breached agricultural security measures and give the phytopathology community the most complete picture of each novel viral sequence.

* Author for correspondence (duffy@sebs.rutgers.edu).

One contribution of 14 to a Theme Issue 'New experimental and theoretical approaches towards the understanding of the emergence of viral infections'.

Table 1. The number of short-form reports in three plant pathology journals that involved sequencing of plant viruses from 2007 to 2009 (2/07–1/10 for *New Disease Reports*).

	short-form reports of plant viruses with		
	PNI ^a	an unseen phylogenetic tree ^b	a phylogenetic tree ^c
<i>New Disease Reports</i>	47	0	5
<i>Plant Disease</i>	154	7	2
<i>Plant Health Progress</i>	9	0	1

^aThose reports that used per cent nucleotide identity (PNI) for any reason, including viral species identification, but no phylogenetic methods.

^bThose reports that mentioned a phylogenetic analysis that was not presented (e.g. 'a maximum-parsimony analysis showed that these sequences group closely together' (Raj *et al.* 2008)).

^cThose reports that provided a phylogenetic tree. The format of *New Disease Reports* and *Plant Health Briefs*, which allow figures as part of the report instead of supplementary 'e-Xtra' information, might make these journals a more welcoming home for reports with phylogenetic trees.

2. MATERIAL AND METHODS

For each analysis, sequences were obtained from GenBank and aligned and trimmed manually with Se-AL v2.0a11 (<http://tree.bio.ed.ac.uk/software>). No statistically significant recombination breakpoints were detected by more than two of the following algorithms as implemented in RDP 3.15 (Martin *et al.* 2005): RDP, GENECONV, Bootscan, MaxChi, Chimaera, SiScan and 3Seq. Therefore, recombination was not considered in further phylogenetic analyses. Nucleotide substitution models were selected by Akaike's information criterion (AIC) with MODELTEST 3.7 (Posada & Crandall 1998). Maximum-likelihood phylogenetic analyses were performed with PAUP* 4.0 beta (Swofford 2003) and bootstrapped with 1000 replicates. Trees were manipulated with FIGTREE v.1.2.3 (<http://tree.bio.ed.ac.uk/software/figtree/>), mid-point rooted for clarity and presented with branch lengths scaled to the numbers of substitutions/site.

3. RESULTS AND DISCUSSION

A survey of several journals that publish short-form reports on new plant diseases or established diseases in new plants or locations revealed that phylogenetic methods are rarely used when describing a novel viral sequence (table 1). Only 3.6 per cent of the more than 200 viral reports published in three journals (BSPP's *New Disease Reports*, APS' *Plant Disease Reports* and *Plant Health Progress*) contained a phylogenetic tree that a reader could look at and evaluate. Far more popular was reporting the PNI to other strains of the virus. Sometimes, the PNI was explicitly aided by NCBI's BLAST (Basic Local Alignment Search Tool), which was used to identify closely related isolates in GenBank for comparison. Often, however, PNI was calculated relative to isolates without any rationale for why the specific isolates were selected. A third category was needed for reports that communicated that they had created a phylogenetic tree, but did not

provide the tree to the reader (though presumably they would, upon individual request). Most of these reports aimed solely to communicate that a virus had been found in a new plant or place, not to say anything about its biogeography, nor assert where the infection had come from. For mere pathogen identification, PNI is adequate.

(a) *Per cent nucleotide identity: good, and sometimes good enough*

If the only goal is determining what virus is present in a diseased plant, then obtaining a sequence with species-specific primers and confirming that it is very similar to known isolates of a virus is sufficient, and is more sensitive than serological techniques (Schneider *et al.* 2004). The vast majority of short-form plant virus reports use sequence data in this way. In fact, some reports do not mention exact PNI values because the authors felt it was sufficient to mention that the sequences were highly similar to one another.

Importantly, many virus families use a threshold PNI to determine if a novel viral sequence represents a new species (determined and revised by the International Committee for Taxonomy of Viruses (Fauquet *et al.* 2005)). For instance, begomoviruses are of the same species if they are more than 89 per cent identical over the full-length DNA-A segments from previously characterized species, while their sister group, the mastreviruses, use a cut-off of 75 per cent (Fauquet *et al.* 2008). The single-stranded RNA potyviruses use a threshold value of 85 per cent (Fauquet *et al.* 2005), but there is discussion of reducing this to 76 per cent (Adams *et al.* 2005). It is necessary to include PNI when characterizing a novel viral sequence from a family that has a threshold PNI in order to assess whether or not it represents a novel species.

(b) *BLAST can be better*

PNI is a better measure than simply the presence or absence of a PCR band, since it confirms that what lies between those sequence-specific sites is the expected sequence, and does not ignore insertions and deletions. If the researchers do not wish to undertake a more complete phylogenetic analysis, using BLAST can be an intermediate step (Altschul *et al.* 1990). BLAST compares a query sequence to the entire GenBank non-redundant nucleotide sequence collection, looks for high identity matches, and selects sequences that closely match the entire query sequence. The matching sequences are ranked by expect scores (E-values) that correspond to the relative likelihood of the match being identified by chance alone. If one uses BLAST to query a novel viral sequence, and it is a member of a viral species or genus that has many sequences in GenBank, the results will show the publicly available sequences of that group to which the novel sequence has the highest PNI. These sequences increasingly have their country and time of isolation in their GenBank files or in an accompanying publication. However, it is important to note that these details must be explicitly specified; the year of submission to GenBank and the country

of the submitting scientists are not reliable indicators of when and where a virus was isolated. By using BLAST, one can find the most similar viruses for PNI comparisons without preconceived notions about sequences from particular geographical areas to which the novel sequence should be most closely related.

(c) *Some situations are perfect for phylogenetics*

If PNI is good, and BLAST is better, then the most thorough placement a novel viral sequence can initially receive is through phylogenetic analysis. Not every new report of a plant disease requires a phylogeny, but reporting the first incidence of a pathogen in a new location, without attempting to determine where it could have come from, shortchanges the scientific community. In order to increase plant biosecurity, each country or region needs to know how and from where pathogens previously entered the region (Rondoni 2009). If novel sequences come with biogeographical information in the initial report, it will speed the process of highlighting frequent sources of infection and consistently leaky ports of entry in interstate and international commerce. These analyses could lead to increased vigilance when screening imports from a subset of countries that are consistent sources of phytopathogens. This is especially critical, as trade agreements have weakened the ability of many countries to routinely quarantine plant material (Jones 2009; Rondoni 2009).

PNI analyses make some of the same implicit assumptions as phylogenetic analyses: that the alignment of the sequences is correct and reflects homology. By choosing one or a few sequences to calculate PNI against, the author is making the *a priori* decision that these are the most informative sequences with which to compare the new sequence. By contrast, phylogenetic analyses that involve more sequences from a wider geographical range than is typically employed in PNI analyses allow unexpected relationships between sequences to emerge.

Recombination can obfuscate patterns of common descent because it means different sequences with different evolutionary histories are physically joined together in the genome. Recombination is a frequent occurrence in many plant viruses (Chare & Holmes 2006; Lefeuvre *et al.* 2007), and software programs exist to detect recombination breakpoints (many are collated into RDP3, (Martin *et al.* 2005)) so that the evolutionary relationships of different portions of the genome can be analysed separately. As PNI makes fewer assumptions about ancestry, it is less affected by the presence of recombination than are phylogenetic analyses. Alignments destined for phylogenies should be screened for recombination, and researchers must be very cautious about further analysis with the entire alignment. One approach is to break up the dataset into smaller alignments at recombination breakpoints and analyse them separately. Another is to aim not for a single, bifurcating phylogenetic tree, but a network (Huson & Bryant 2006). The blended evolutionary history of several plant viruses has been traced using split network methods in SPLITSTREES4

(Hu *et al.* 2007; Codoner & Elena 2008; Martin *et al.* 2009). While networks can better reflect the true ancestry of many recombinant plant viruses, the analysis of migration and hypothesis testing is more complicated on a network than bifurcating trees (Huson & Bryant 2006).

BLAST analysis shares many assumptions with phylogenetic analyses, and has an understandable bias towards finding the most closely related sequence over the longest stretch of nucleotides. BLAST will score a longer but lower similarity match higher than a shorter, more similar match. As many viral sequences in GenBank are from relatively short species-specific PCR amplicons, this means that using BLAST on a longer viral sequence, such as the whole genome, will not necessarily return the highest PNI match over the highly sequenced, species-specific amplicon region. Rather, it will return the highest PNI match for the entire length of the query sequence. Alignments for phylogenetic analyses can be trimmed such that all sequences have the same length and the algorithms compare across an even amount of information. There is an obvious disadvantage to eliminating part of the new viral sequence from analysis and consideration, but perhaps the solution is to present both a phylogenetic analysis based on a good alignment and the PNI from a BLAST analysis with the full sequence.

Phylogeographical approaches not only allow epidemiologists to trace the source of infections, they also can provide a measure of how confident researchers should be in their conclusions. Through estimates of support for clades, from bootstrap analyses or from Bayesian posterior probabilities, authors and readers alike can assess how probable it is that the alignment underlying the phylogenetic tree is representative of a real relationship between sequences that are closely grouped on the tree. This allows authors to move beyond suggestion and give a relative measure of the support for their assertions. For instance, when zucchini yellow mosaic virus (ZYMV) was first discovered in Poland, pair-wise PNI comparisons indicated it was more closely related to sequences from Asia than to other European strains (Pospieszny *et al.* 2007). However, these authors were unable to give a measure of how confident they were in these relationships. The phylogenetic relationship between Polish (and now two French) ZYMV isolates and a Chinese isolate from 1999 was published in 2009, and revealed moderate 77 per cent bootstrap support, lending increased credibility to the authors' conclusions (Lecoq *et al.* 2009). Several additional examples of the utility of phylogenetic analysis inspired by the recent plant virus report literature are given below.

(d) *Tomato yellow leaf curl virus*

As TYLCV has continued to spread in North America and the Caribbean, sequences have recently been added to GenBank from viruses isolated in Arizona (Idris *et al.* 2007), Guadeloupe, Grenada, Kentucky (de Sá *et al.* 2008), Martinique (Urbino & Dalmon 2007) and Mexico (Idris *et al.* 2007; Gámez-Jiménez *et al.* 2009). An updated tree created from an

Figure 1. Maximum-likelihood tree of a 488-base alignment of partial coat protein sequences of New World isolates of tomato yellow leaf curl virus with three non-North American isolates, constructed with a Tamura–Nei (TrN) model of nucleotide substitution. Branches with greater than or equal to 75% bootstrap support are labelled. Taxon labels are as previously published (Duffy & Holmes 2007) or labelled according to convention (Fauquet *et al.* 2008) and given with its GenBank accession number.

alignment of partial coat protein genes, which recapitulates the two geographically distinct New World clades (Duffy & Holmes 2007; Zhang *et al.* 2009), is given in figure 1. The more recently isolated viruses were published with PNI to a range of closely related TYLCV isolates, but our phylogenetic analysis offers better resolution of the ancestry of some of these strains. The publication describing the partial genome sequences of TYLCV isolated in Kentucky reported them to be 98–99% identical to TYLCV-US:TX, TYLCV-US:AZ and TYLCV-US:SC (de Sá *et al.* 2008), but our analysis provides support for the Kentucky strains being more closely related to the Texan strain in particular. Similarly, the first Californian TYLCV isolate had a very high PNI to the strains from western Mexico (Rojas *et al.* 2007), but the tree in figure 1 places a quantitative measure of the level of support in the alignment for the grouping of TYLCV-US:CA with the Sonoran and Sinaloan strains.

(e) *Banana bunchy top virus*

The economically important banana bunchy top virus (BBTV) is a threat to banana crops in Asia, on Pacific

Islands, in Australia and in the Middle East (Amin *et al.* 2008). The spread of BBTV into several novel geographical areas has been documented, especially in Hawaii, where the introduction of the virus to each island can be traced and dated using molecular phylogenetic methods (Almeida *et al.* 2009). That analysis revealed evidence of two introductions of BBTV onto Kauai island, despite 99.6 per cent or more nucleotide identity among the Kauaian sequences. Thus, the use of phylogenetic methods revealed continued exchange of infected plant material and/or infected banana aphids among the Hawaiian islands, which would have almost certainly been overlooked if the researchers only used PNI (Almeida *et al.* 2009).

We constructed a gene genealogy for all available coat protein genes of BBTV to see if we could detect any other cryptic second introductions of this virus. In figure 2, we highlight sequences from the Hainan province in China and show that two coat protein alleles are circulating within the region. Hainan is an island in the South China Sea and its first BBTV isolates grouped with those of nearby Vietnam (Jun & Zhi-Xin 2005). The more recently sequenced isolate from June 2008 is more closely related to viruses

Figure 2. Maximum-likelihood tree of a 513-base alignment of coat protein sequences (the 513-base ORF on DNA-3) of banana bunchy top virus, constructed with a TrN substitution model including the fraction of invariant sites (I) and substitution rate heterogeneity among sites (T). Branches with greater than or equal to 75% bootstrap support are labelled. Taxon labels are GenBank accession numbers preceded by two-letter country codes: AU, Australia; BI, Burundi; CM, Cameroon; CN, China; EG, Egypt; FJ, Fiji; ID, Indonesia; IN, India; JP, Japan; MM, Myanmar; PH, Philippines; PK, Pakistan; TW, Taiwan; VN, Vietnam. Isolates from Hainan, China are shown in black, other locations are shown in grey.

from the Chinese mainland (direct submission to GenBank, accession number FJ463044). The older Hainan BBTV isolates are 99.69 and 99.57 per cent identical to FJ463044—again very high PNIs that make an incorrect, direct, ancestral relationship between the older and newer Hainan isolates seem plausible.

(f) *Tomato chlorosis virus*

In 2006–2007 tomato chlorosis virus (ToCV) was identified for the first time in South America, in Sumaré, Brazil (Barbosa *et al.* 2008). The researchers who sequenced portions of this Brazilian ToCV found that the strain was 99 per cent identical to ToCV from

the USA. The phylogenetic analysis in figure 3 suggests, albeit with lower bootstrap support than that observed in our other analyses, that ToCV in Brazil is more closely related to ToCV from the Mediterranean (Greece, Turkey and Lebanon) than to strains from North and Central America. When the 463 base partial genome sequence was trimmed to make it align with other ToCV sequences in GenBank, the 309 base region of the Brazilian isolate was more than 99 per cent identical to Turkish and Lebanese isolates, and 98.7 per cent or less identical to the other New World isolates. In addition to the lower PNI in this region than what the Brazilian strain shares with the Mediterranean strains, the Brazilian strain does not have two common synapomorphies

Figure 3. Maximum-likelihood tree of a 309-base alignment of partial heat shock protein 70 sequences of tomato chlorosis virus, constructed with a transversional (TVM) model of nucleotide substitution. Branches with greater than or equal to 50% bootstrap support are labelled. Taxon labels are GenBank accession numbers preceded by two-letter country codes: BR, Brazil; CU, Cuba; ES, Spain; FR, France; GR, Greece; IT, Italy; LB, Lebanon; MX, Mexico; MU, Mauritius; PT, Portugal; RE, Reunion Island; TR, Turkey; YT, Mayotte.

(mutations) shared among the other New World isolates. This analysis suggests that ToCV might have been introduced twice to the New World, but the relatively weak bootstrap values on the tree make any definitive statement inadvisable. As future isolates are collected and analysed, the hypothesis of multiple introductions can be more thoroughly examined.

These examples show that a phylogenetic approach can provide a geographical context to novel viral sequences, and either provide support for intuitive relationships, or introduce the notion that viruses have migrated multiple times into a region.

(g) No additional experiments

For researchers who are already using sequencing to identify and confirm the causative agent of a disease, no additional wet-lab work is needed to conduct a phylogenetic analysis. While there is a learning curve for using phylogenetic programs, many of the relevant programs are free or low-cost, and tutorials exist online and in book form. One approachable volume that now assists readers using MEGA (Tamura *et al.* 2007) is Barry Hall's third edition of *Phylogenetic Trees Made Easy* (Hall 2007). Beginning to use phylogenetic methods opens the door to more advanced hypothesis testing. One directly relevant application would be comparing the likelihood of two hypothetical evolutionary histories: one where a virus is allowed to

have multiple introductions to a geographical region, and one where all isolates from the geographical region must be descended from a single introduction (e.g. Duffy & Holmes 2007).

In addition to noting multiple introductions of a virus and identifying weak points in plant biosecurity, it is important for disease management to know that multiple strains of a virus are circulating in the same region. Co-infection of the same plant by multiple strains of a virus can lead to more severe symptoms owing to synergistic action or the rare creation of a more virulent genotype, both illustrated in the cassava mosaic disease outbreak in Uganda in the late 1990s (Legg *et al.* 2006). With foreknowledge of multiple strains in an area, researchers could begin monitoring to see if recombinant strains emerge and are associated with more severe symptoms. However, plant pathologists need to be aware of the potential problem before they put the time and resources into increased vigilance.

This work was supported by funds from the Rutgers School of Environmental and Biological Sciences and the New Jersey Agricultural Experiment Station.

REFERENCES

- Adams, M. J., Antoniwi, J. F. & Fauquet, C. M. 2005 Molecular criteria for genus and species discrimination within the family *Potyviridae*. *Arch. Virol.* **150**, 459–479. (doi:10.1007/s00705-004-0440-6)
- Akad, F., Jacobi, J. C. & Polston, J. E. 2007 Identification of *Tomato yellow leaf curl virus* and *Tomato mottle virus* in two counties in Alabama. *Plant Dis.* **91**, 906. (doi:10.1094/PDIS-91-7-0906C)
- Almeida, R. P., Bennett, G. M., Anhalt, M. D., Tsai, C. W. & O'Grady, P. 2009 Spread of an introduced vector-borne banana virus in Hawaii. *Mol. Ecol.* **18**, 136–146. (doi:10.1111/j.1365-294X.2008.04009.x)
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
- Amin, I., Qazi, J., Mansoor, S., Ilyas, M. & Briddon, R. W. 2008 Molecular characterization of *Banana bunchy top virus* (BBTV) from Pakistan. *Virus Genes* **36**, 191–198. (doi:10.1007/s11262-007-0168-y)
- Barbosa, J. C., Teixeira, A. P. M., Moreira, A. G., Camargo, L. E. A., Bergamin Filho, A., Kitajima, E. W. & Rezende, J. A. M. 2008 First report of *Tomato chlorosis virus* infecting tomato crops in Brazil. *Plant Dis.* **92**, 1709. (doi:10.1094/PDIS-92-12-1709C)
- Brown, J. K. & Idris, A. M. 2006 Introduction of the exotic monopartite *Tomato yellow leaf curl virus* into west coast Mexico. *Plant Dis.* **90**, 1360. (doi:10.1094/PD-90-1360A)
- Chare, E. R. & Holmes, E. C. 2006 A phylogenetic survey of recombination frequency in plant RNA viruses. *Arch. Virol.* **151**, 933–946. (doi:10.1007/s00705-005-0675-x)
- Codoner, F. M. & Elena, S. F. 2008 The promiscuous evolutionary history of the family *Bromoviridae*. *J. Gen. Virol.* **89**, 1739–1747. (doi:10.1099/vir.0.2008/000166-0)
- de Sá, P. B., Seebold, K. W. & Vincelli, P. 2008 First report of *Tomato yellow leaf curl virus* in greenhouse tomatoes in Kentucky. *Plant Health Prog.* (doi:10.1094/PHP-2008-0819-01-RS)
- Duffy, S. & Holmes, E. C. 2007 Multiple introductions of the Old World begomovirus *Tomato yellow leaf curl virus* into the New World. *Appl. Environ. Microbiol.* **73**, 7114–7117. (doi:10.1128/AEM.01150-07)

- Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U. & Ball, L. A. 2005 *Virus Taxonomy: VIIIth Report of the Int. Committee on Taxonomy of Viruses*. San Diego, CA: Elsevier Academic Press.
- Fauquet, C. M., Briddon, R. W., Brown, J. K., Moriones, E., Stanley, J., Zerbini, M. & Zhou, X. 2008 Geminivirus strain demarcation and nomenclature. *Arch. Virol.* **153**, 783–821. (doi:10.1007/s00705-008-0037-6)
- Gómez-Jiménez, C., Romero-Romero, J. L., Santos-Cervantes, M. E., Leyva-López, N. E. & Méndez-Lozano, J. 2009 Tomatillo (*Physalis ixocarpa*) as a natural new host for *Tomato yellow leaf curl virus* in Sinaloa, Mexico. *Plant Dis.* **93**, 545. (doi:10.1094/PDIS-93-5-0545A)
- Hall, B. G. 2007 *Phylogenetic trees made easy: a how-to manual*. Sunderland, MA: Sinauer Associates.
- Hu, J. M., Fu, H. C., Lin, C. H., Su, H. J. & Yeh, H. H. 2007 Reassortment and concerted evolution in *Banana bunchy top virus* genomes. *J. Virol.* **81**, 1746–1761. (doi:10.1128/JVI.01390-06)
- Huson, D. H. & Bryant, D. 2006 Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267. (doi:10.1093/molbev/msj030)
- Idris, A. M., Guerrero, J. C. & Brown, J. K. 2007 Two distinct isolates of *Tomato yellow leaf curl virus* threaten tomato production in Arizona and Sonora, Mexico. *Plant Dis.* **91**, 910. (doi:10.1094/PDIS-91-7-0910C)
- Isakeit, T., Idris, A. M., Sunter, G., Black, M. C. & Brown, J. K. 2007 *Tomato yellow leaf curl virus* in tomato in Texas, originating from transplant facilities. *Plant Dis.* **91**, 466. (doi:10.1094/PDIS-91-4-0466A)
- Jones, R. A. 2009 Plant virus emergence and evolution: origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. *Virus Res.* **141**, 113–130. (doi:10.1016/j.virusres.2008.07.028)
- Jun, T. E. Z. & Zhi-Xin, L. 2005 Cloning and sequencing of DNA components of Banana bunchy top virus Hainan isolate. *Chin. J. Agric. Biotechnol.* **2**, 91–97. (doi:10.1079/CJB200566)
- Lecoq, H., Wipf-Scheibel, C., Chandeysson, C., Le Van, A., Fabre, F. & Desbiez, C. 2009 Molecular epidemiology of *Zucchini yellow mosaic virus* in France: an historical overview. *Virus Res.* **141**, 190–200. (doi:10.1016/j.virusres.2008.11.020)
- Lefeuvre, P., Lett, J. M., Reynaud, B. & Martin, D. P. 2007 Avoidance of protein fold disruption in natural virus recombinants. *PLoS Pathog.* **3**, e181. (doi:10.1371/journal.ppat.0030181)
- Legg, J. P., Owor, B. E., Sseruwagi, P. & Ndunguru, J. 2006 Cassava mosaic virus disease in East and Central Africa: epidemiology and management of a regional epidemic. In *Plant virus epidemiology* (ed. J. M. Thresh). San Diego, CA: Academic Press.
- Martin, D. P., Williamson, C. & Posada, D. 2005 RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**, 260–262. (doi:10.1093/bioinformatics/bth490)
- Martin, S., Sambade, A., Rubio, L., Vives, M. C., Moya, P., Guerri, J., Elena, S. F. & Moreno, P. 2009 Contribution of recombination and selection to molecular evolution of *Citrus tristeza virus*. *J. Gen. Virol.* **90**, 1527–1538. (doi:10.1099/vir.0.008193-0)
- Moury, B., Desbiez, C., Jacquemond, M. & Lecoq, H. 2006 Genetic diversity of plant virus populations: towards hypothesis testing in molecular epidemiology. In *Plant virus epidemiology* (ed. J. M. Thresh). San Diego, CA: Academic Press.
- Ndunguru, J., Legg, J. P., Aveling, T. A. S., Thompson, G. & Fauquet, C. M. 2005 Molecular biodiversity of cassava begomoviruses in Tanzania: evolution of cassava geminiviruses in Africa and evidence for East Africa being a center of diversity of cassava geminiviruses. *Virol. J.* **2**, 21. (doi:10.1186/1743-422X-2-21)
- Nischwitz, C., Pappu, H. R., Mullis, S. W., Sparks, A. N., Langston, D. R., Csinos, A. S. & Gitaitis, R. D. 2007 Phylogenetic analysis of *Iris yellow spot virus* isolates from onion (*Allium cepa*) in Georgia (USA) and Peru. *J. Phytopathol.* **155**, 531–535. (doi:10.1111/j.1439-0434.2007.01272.x)
- Posada, D. & Crandall, K. A. 1998 MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818. (doi:10.1093/bioinformatics/14.9.817)
- Pospieszny, H., Hasiow, B. & Borodynko, N. 2007 A Polish isolate of *Zucchini yellow mosaic virus* from zucchini is distinct from other European isolates. *Plant Dis.* **91**, 639–639. (doi:10.1094/PDIS-91-5-0639B)
- Raj, S. K., Kumar, S., Snehi, S. K. & Pathre, U. 2008 First report of *Cucumber mosaic virus* on *Jatropha curcas* in India. *Plant Dis.* **92**, 171. (doi:10.1094/PDIS-92-1-0171C)
- Rojas, M. R., Kon, T., Natwick, E. T., Polston, J. E., Akad, F. & Gilbertson, R. L. 2007 First report of *Tomato yellow leaf curl virus* associated with tomato yellow leaf curl disease in California. *Plant Dis.* **91**, 1056. (doi:10.1094/PDIS-91-8-1056A)
- Rondoni, B. 2009 The role of plant biosecurity in preventing and controlling emerging plant virus disease epidemics. *Virus Res.* **141**, 150–157.
- Schneider, W. L., Sherman, D. J., Stone, A. L., Damsteegt, V. D. & Frederick, R. D. 2004 Specific detection and quantification of *Plum pox virus* by real-time fluorescent reverse transcription-PCR. *J. Virol. Methods* **120**, 97–105. (doi:10.1016/j.jviromet.2004.04.010)
- Swofford, D. L. 2003 *PAUP* Phylogenetic analysis using parsimony (*and other methods)*. Sunderland, MA: Sinauer Associates.
- Tamura, K., Dudley, J., Nei, M. & Kumar, S. 2007 MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599. (doi:10.1093/molbev/msm092)
- Urbino, C. & Dalmon, A. 2007 Occurrence of *Tomato yellow leaf curl virus* in tomato in Martinique, Lesser Antilles. *Plant Dis.* **91**, 1058. (doi:10.1094/PDIS-91-8-1058A)
- Varsani, A. et al. 2009 Comparative analysis of *Panicum streak virus* and *Maize streak virus* diversity, recombination patterns and phylogeography. *Virol. J.* **6**, 194. (doi:10.1186/1743-422X-6-194)
- Zambrano, K., Carbaloo, O., Geraud, F., Chirinos, D., Fernández, C. & Marys, E. 2007 First report of *Tomato yellow leaf curl virus* in Venezuela. *Plant Dis.* **91**, 768. (doi:10.1094/PDIS-91-6-0768A)
- Zhang, H., Gong, H. & Zhou, X. 2009 Molecular characterization and pathogenicity of *Tomato yellow leaf curl virus* in China. *Virus Genes* **39**, 249–255. (doi:10.1007/s11262-009-0384-8)