Review

Time – the emerging dimension of plant virus studies

A. J. Gibbs,[1] D. Fargette,[2] F. García-Arenal[3] and M. J. Gibbs[4]

Correspondence
A. J. Gibbs
adrian_j_gibbs@hotmail.com

[1]7 Hutt Street, Yarralumla, ACT 2600, Australia

[2]IRD, UMR 186/RPB 'Resistance des Plantes aux Bioagresseurs', BP 64501, 34394 Montpellier cedex 5, France

[3]Centro de Biotecnología y Genómica de Plantas (UPM-INIA) and E.T.S.I. Agrónomos, Campus Montegancedo, Universidad Politécnica de Madrid, Pozuelo de Alarcón, 28223 Madrid, Spain

[4]79 Carruthers Street, Curtin, ACT 2605, Australia

Recent research has revealed that some plant viruses, like many animal viruses, have measurably evolving populations. Most of these viruses have single-stranded positive-sense RNA genomes, but a few have single-stranded DNA genomes. The studies show that extant populations of these viral species are only decades to centuries old. The genera in which they are placed have diverged since agriculture was invented and spread around the world during the Holocene period. We suggest that this is not mere coincidence but evidence that the conditions generated by agriculture during this era have favoured particular viruses. There is also evidence, albeit less certain, that some plant viruses, including a few shown to have measurably evolving populations, have much more ancient origins. We discuss the possible reasons for this clear discordance between short- and long-term evolutionary rate estimates and how it might result from a large timescale dependence of the evolutionary rates. We also discuss briefly why it is useful to know the rates of evolution of plant viruses.

Introduction

In 2002, Stenger *et al.* (2002) reported that differences that appeared during passaging of wheat streak mosaic virus (WSMV), taken together with differences within its population in North America, indicated that this virus had been diverging at about 1.1×10^{-4} nucleotide substitutions (ns) per site per year (Stenger *et al.*, 2002) since it first invaded those wheat crops in the early 20th century (McKinney, 1937). This was notable as the first of several reports showing that populations of some plant viruses evolve quickly enough for their evolution to be directly observed. They have measurably evolving populations (Drummond *et al.*, 2003) and their mutation rates are similar to those reported for some animal viruses and bacteriophages with RNA genomes (Drake *et al.*, 1998; Duffy *et al.*, 2008; Fargette *et al.*, 2008a; Sanjuán *et al.*, 2009), such as influenza virus and human immunodeficiency virus (HIV) (Hanada *et al.*, 2004; Jenkins *et al.*, 2002). These records contrast with other earlier reports suggesting that plant viruses may evolve more slowly than animal viruses (Gibbs *et al.*, 2008a).

Here, we review and discuss these reports and some of the interesting matters arising.

The evidence

The evolutionary rates of organisms with measurably evolving populations can be estimated most directly by comparing samples collected serially. However, in practice, evolutionary rates are estimated most often from phylogenetic analysis of gene sequences obtained from natural populations after the removal of recombinant sequences. Recombination (and reassortment) between genomes confounds attempts to estimate evolutionary rates, but is readily detected (Fourment *et al.*, 2008; Martin *et al.*, 2005). If the samples providing the sequences were collected on known dates over a sufficient period of time for evolutionary changes to have occurred, a strategy called heterochronous sampling (Drummond *et al.*, 2003), then the evolutionary rate(s) can be determined directly by comparing the 'tip dates' of their phylogenies using regression and statistical inference methods. If, however, a population has been sampled on only a single occasion or over such a short period that no evolution has occurred, then its evolutionary rate can only be estimated if the dates of nodes within its phylogeny are known from historical or other data. Finally, there is the possibility, when viral (or other pathogen) and host phylogenies are found to be congruent, that the viruses and their hosts have codiverged and so the rates and datings for the hosts, obtained from fossils or other evidence, can be used for the viruses.

The evolutionary rates of different plant viruses have been estimated using each of these methods (see Table 1 for rates and references), as described below.

Serial sampling has been used to estimate the rates of evolution of banana bunchy top virus (BBTV) and maize

Table 1. Estimated rates of plant virus evolution

Virus	ns per site per year	Source
Serial sampling		
BBTV	3.9×10^{-4}	Almeida *et al.* (2009)
MSV-1	$7.4–7.9 \times 10^{-4}$	van der Walt *et al.* (2008)
MSV-2	$2–3 \times 10^{-4}$	Harkins *et al.* (2009)
Heterochronous sampling		
BBTV	1.4×10^{-4}	Almeida *et al.* (2009)
Barley yellow dwarf viruses	$4.8–6.3 \times 10^{-4}$	Malmstrom *et al.* (2007)
East African cassava mosaic begomovirus	$1.6–1.33 \times 10^{-4}$	Duffy & Holmes (2009)
RYMV	48×10^{-4}	Fargette *et al.* (2008a)
TYLCVs (eight species)	2.9×10^{-4}	Duffy & Holmes (2008)
ZYMV	5.0×10^{-4}	Simmons *et al.* (2008)
Node dating		
Potyviruses (20 species)	1.15×10^{-4}	Gibbs *et al.* (2008c)
WSMV	1.1×10^{-4}	Stenger *et al.* (2002)
Codivergence studies		
Cereal mastreviruses	10^{-8}	Wu *et al.* (2008)
TYMV	$1.3 \times 10^{-6–7}$	Blok *et al.* (1987); Gibbs *et al.* (1986); Guy & Gibbs (1981)
Tobamoviruses*	2.2×10^{-8}	Gibbs (1980); Gibbs *et al.* (2008a)
South American begomoviruses	$<0.6 \times 10^{-6}$	This review

*Calculated from a maximum-likelihood analysis of 320 tobamovirus CP genes representing around 25 species (Gibbs *et al.*, 2008a). The sequences linked through the basal node of the phylogeny differed by 2.159 ($\pm$0.206) ns per site, and it is assumed that these viruses first diverged 100 million years ago.

streak virus (MSV). Studies of this sort may have errors when the initial inocula have not been cloned, or if unusual hosts and methods for transmitting between them have been used. Nonetheless the reported rates (Table 1) have been found to be closely similar to those obtained by heterochronous sampling.

Heterochronous sampling has been used in studies of populations of BBTV, barley yellow dwarf viruses (BYDVs), rice yellow mottle virus (RYMV), tomato yellow leaf curl virus (TYLCV) and zucchini yellow mosaic virus (ZYMV). All were found to be evolving at measurable rates (Table 1).

Dated nodes were used to estimate the rates of evolution of WSMV in North American wheat crops and for a sample of 20 potyvirus species. These studies (Table 1) illustrate how difficult it is to obtain reliable dates for nodes; nonetheless the rate estimates were close to those obtained by serial or heterochronous sampling.

Codivergence was inferred for cereal dwarf mastreviruses and their monocotyledonous hosts. This suggested that as cereals diverged at around 25 million years ago, cereal dwarf mastreviruses did so too, indicating that mastreviruses have a long-term evolutionary rate of 10^{-8} ns per site per year (Table 1). Likewise, the broad similarity between the sequence-based phylogeny of the tobamoviruses and the phylogeny of their preferred angiosperm host plants suggests that they codiverged; as the angios-

perms are 140–180 million years old (Soltis *et al.*, 2008), this suggests that the tobamoviruses have a long-term evolutionary rate close to that of the mastreviruses (Table 1).

The discovery of viral gene fragments in the genomes of likely hosts is also emerging as a source of evidence of ancient populations of viruses. Most conclusive of these is the observation (Ashby *et al.*, 1997; Bejarano *et al.*, 1996; Murad *et al.*, 2004) of repetitive geminivirus-related DNA sequences (GRDs) in the genome of tobacco (*Nicotiana tabacum*), an allotetraploid, and its paternal diploid ancestor *N. tomentosiformis*, and also in the sister species of that ancestor, *N. tomentosa* and *N. kawakamii*. GRDs were not found in *N. otophora*, the other more distant *Nicotiana* species in the section Tomentosae (Fig. 1), nor in eight other more distantly related *Nicotiana* species and six other solanaceous and non-solanaceous species. The GRDs are related to the *rep* and *ori* genes of geminiviruses and fall into two families with distinct, but closely related, sequences: the GRD5 family occurs in homologous chromosome 4 of *N. tabacum* and its three diploid relatives, whereas the GRD3 family only occurs in *N. tabacum* and *N. tomentosiformis*, suggesting that ancestral GRDs successfully integrated with a *Nicotiana* genome on two occasions. BLAST searches of the international gene sequence databases (done in January 2009) suggest that the geminivirus progenitor of the GRDs was either a begomovirus or perhaps a curtovirus, but not a
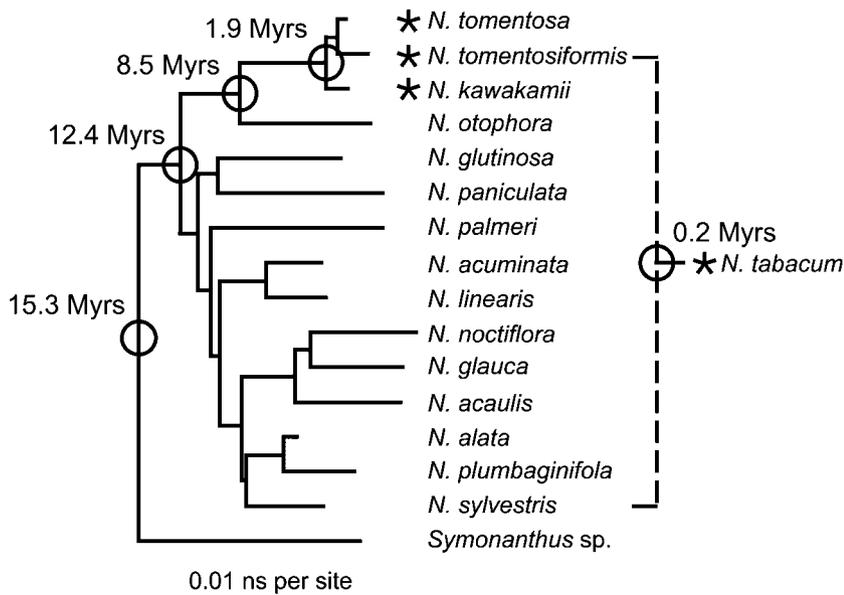
Fig. 1. Phylogenetic relationships of representative diploid *Nicotiana* species and the allotetraploid *N. tabacum* inferred by comparing concatenated sequences of *matK*, *ndhF*, *trnL* intron, *trnL-F* spacer and *trnS-G* spacer (5260 nt in each concatenate) (Clarkson *et al.*, 2005). The phylogenetic relationships of the sequences were inferred and compared using the maximum-likelihood method PhyML (Guindon & Gascuel, 2003) with the general time-reversible model with gamma-distributed rate variation and a proportion of invariable sites. All other optional parameters were estimated by the program. The resulting trees were converted to patristic distances (Fourment & Gibbs, 2006) and the dates of the nodes were estimated assuming that *Nicotiana* and *Symonanthus* diverged 15.3 million years ago (Clarkson *et al.*, 2005). Species marked with an asterisk contained repetitive geminivirus-related DNA sequences. They have not been found in other *Nicotiana* spp. nor in more distantly related solanaceous species. The broken line indicates the diploid parents from which *N. tabacum* originated within the last 0.2 million years. Myrs, million years.

mastrevirus, as the GRDs match the gene encoding the replication initiator (replication-associated) proteins (C1 or AC1) of the begomoviruses and the curtoviruses. Curtoviruses are recombinants that acquired their C1/Rep ORF from a begomovirus (Rybicki, 1994). Curtoviruses only match the GRD sequences when GenBank is searched with the nucleotide sequences (BLASTN) not the encoded amino acid sequences (BLASTX), which suggests that biases have arisen in the GRD gene sequences since they became genomic and these differences are not translated.

The likely timing of the GRD integration events can be estimated from the phylogeny of the genus *Nicotiana*. The genera *Nicotiana* and *Symonanthus* diverged about 15.3 million years ago, and *N. tabacum*, an allopolyploid originated from diploid parents within the last 0.2 million years (Clarkson *et al.*, 2004, 2005). A phylogeny of 15 diploid *Nicotiana* species shows that diploid *Nicotiana* species containing GRD5 are monophyletic and diverged from one another 1.9 million years ago (Fig. 1). Therefore, we conclude that, as all the species of *Nicotiana* containing GRDs were confined to South America until five centuries ago, it is probable that there were begomoviruses in South America more than 1.9 million years ago. A maximum-likelihood analysis of the aligned sequences of the core coat protein (CP) genes of 98 South American begomoviruses shows that they have a maximum difference of approximately 1.1 ns per site. Therefore, the long-term rate of

evolution of the CP gene in South American begomoviruses is no greater than 0.6×10^{-6} ns per site per year.

Other fragments of evidence indicate that the tobamoviruses evolve slowly. One is from herbarium specimens of *N. glauca*, a species that entered Australia in the late 19th century. The specimens were up to 100 years old and some were found to contain either tobacco mosaic virus (TMV) or tobacco mild green mosaic virus (TMGMV) or both. Two regions of the genomes of these isolates were sequenced. The sequences of TMGMV, but not those of TMV, showed no time-related changes (Fraile *et al.*, 1997), suggesting that no divergence had occurred over 100 years. These results are congruent with the small mutation rates reported for TMGMV in passage experiments (Rodriguez-Cerezo & García-Arenal, 1989) and the limited genetic diversity of TMGMV isolates from *N. glauca* from four continents (Fraile *et al.*, 1996).

Another line of evidence indicating that tobamoviruses are ancient was reported by Holmes (1951) who noted that the species of *Nicotiana* that respond to infection by TMV in a hypersensitive manner are all natives of Central and South America: *N. glutinosa* is a native of Peru, *N. repanda* of Mexico, *N. rustica* of Ecuador and Peru, and *N. langsdorfii* of Brazil. Several species of other genera of the Solanaceae that are native to South America behave in the same way, including *Solanum capsicastrum* of Brazil and *S. tuberosum* of Bolivia and Peru. In contrast, *Nicotiana* species that are

most susceptible to systemic TMV infection, and produce the greatest concentrations of virions, are mostly found in North America, southern South America and Australia. Holmes argued that these differences 'would seem to imply that the original habitat of tobacco mosaic virus was within an area of the New World, centering about some part of Peru, Bolivia, or Brazil' (Holmes, 1951), and noted that *N. tabacum* itself is a species found only in crops or as a crop fugitive, and is unlikely to have been the original long-term host of TMV.

Another virus that has provided some evidence of a sluggish long-term evolutionary rate is turnip yellow mosaic virus (TYMV). Two distinct lineages of the virus are widespread in Europe, and one of them is also in Japan (Kirino *et al.*, 2008) and in the Kosciusko alpine area of south-eastern Australia (Guy & Gibbs, 1981). CP gene sequences of the three populations (Blok *et al.*, 1987; Hayden *et al.*, 1998; Kirino *et al.*, 2008) form separate clusters that are approximately equidistant (0.063–0.078 ns per site in trees calculated as in Fig. 1). Biological, geographical, palynological and palaeoclimatological data suggest that TYMV entered Australia before the last glaciation of the region (12 000 years ago) but not before the last major interglacial climatic period 125 000 years ago (Gibbs *et al.*, 1986, 1989; Ruddiman, 2003). As the isolates connected through the basal node of the Australian TYMV CP gene phylogeny have diverged by 0.031 ns per site, it seems that the long-term evolutionary rate of that population is 1.3×10^{-6}–1.3×10^{-7} ns per site per year (Table 1).

Viral origins, evolution and the invention of agriculture

Two important conclusions emerge from the above synopsis of the rates of evolution of plant viruses. Firstly, many populations of plant viruses are evolving at measurable rates, indeed so fast that the most recent common ancestor (TMRCA) of each the present populations first replicated only a few centuries or millennia ago. Secondly, the evolutionary rates estimated from co-divergence studies suggest that some plant viruses have much more ancient origins but, perplexingly, a few of these viruses are among those with measurably evolving populations.

The evolutionary rate estimates obtained from serial or heterochronous sampling, or from node dating, suggest that extant species all arose in recent times and at least three of the most damaging plant virus genera radiated within the last few thousand years: luteovirids (i.e. species of the Luteoviridae), 9000 years ago; potyviruses, 6600 years ago; sobemoviruses, 3000 years ago; and the TYLCVs, which are probably representative of the begomoviruses, 10 000 years ago. Thus, these major taxa of crop-infecting viruses all diversified in the period since humankind invented agriculture (Fargette *et al.*, 2008b; Gibbs *et al.*, 2008c) although, of course, they may have originated even earlier and survived as small founder populations. This

raises the question as to whether the invention and spread of agriculture itself triggered and fostered their radiation and dominance. Agriculture was invented by humans 8000–13 000 years ago in at least nine separate regions of the world. In each region, a different set of plant and animal species was domesticated (Bellwood, 2005; Murphy, 2007; Vavilov, 1940). So, what was the worldwide stimulus for this major innovation given that modern hunter-gatherer humans had appeared a long time before? The most likely explanation is that the Holocene period provided a more stable and consistent pattern of world climates than the preceding 250 000 years of the Pleistocene era (Burroughs, 2005), and permitted the slow process of selection and genetic adaptation of domesticated strains of animals and plants to proceed. During their domestication, all plant species, except three, only spread to contiguous regions, but after long distance marine trade was established five centuries ago, most species quickly spread worldwide; the three exceptions are coconut (*Cocus nucifera*), calabash gourd (*Lagenaria siceraria*) and sweet potato or kumara (*Ipomoea batatus*).

One likely consequence of the spread of agriculture has been the greatly increased opportunity for novel encounters between wild and cultivated plant species and their pathogens and vectors (Jones, 2009). These new encounters will have favoured the selection and emergence of plant viruses suited to the new conditions. Crowding of plants associated with agriculture, and especially monocultures of chosen species, facilitated the build-up of vector populations and the spread of pathogens. Crops will foster viruses with particular ecological lifestyles. Crops provide vectors with a predictable succession of suitable herbaceous hosts on which they can reproduce rapidly and produce very large migrant populations, and virus infections often aid vector infestations and virus spread (Baker, 1960; Jiu *et al.*, 2007). Thus, since the mid-Holocene, agriculture may have fostered the emergence, spread and dominance of virus taxa with particular ecologies suited to life in the continuously disturbed environments provided by humans. Viruses seem to be much more widespread in crop and weed species in India, south-east Asia and, to a lesser extent, Europe than in Australia and the Americas; this distribution possibly correlates with the length of time that agriculture has been established in those regions. A link between pathogen emergence and agriculture has also been suggested for fungal and bacterial plant pathogens (Stukenbrock & McDonald, 2008), and it is clear that, in the same way, the Neolithic age increased contact between humans and other animals and fostered the emergence of infectious human diseases (Wolfe *et al.*, 2007). Furthermore, the spread of all crop pathogens must have been greatly increased when marine trade developed and cultivated plant species were carried around the world (Harlan, 1976) together with their weeds (Crosby, 2004), pathogens and parasites.

Short-term evolutionary rate estimates also show that the extant populations of different viral species are only

decades to centuries old. For example, a comparison of gene sequences of 253 RYMV samples collected hetero-chronously between 1996 and 2006 from all over Africa showed clearly that they are a single population which diversified approximately 200 years ago, centuries after African rice, *Oryza glaberrima*, was domesticated or Asian rice, *Oryza sativa*, was introduced (Sweeney & McCouch, 2007), and decades before epidemics of RYMV were reported. Similar studies of the gene sequences of around 40 Australian potyviruses showed that around half of these, found only in crops, have arrived in Australia from overseas since Europeans colonized the continent two centuries ago. The others probably arrived two millennia ago, perhaps transported by Austronesians in their crop plants as they colonized the islands of the Pacific Ocean (Gibbs *et al.*, 2008b, c, d).

Evolutionary rates: the dilemma

The fast rates of evolution found for some populations of viruses contrast with the much slower rates found for others. Significantly, some plant viruses are in both categories. However, the discordance between short- and long-term rates is not confined to viruses of plants. The populations of human immunodeficiency viruses have TMRCAs that suggest that they are only centuries old (Wertheim & Worobey, 2009). However, the taxonomy of endogenous retrovirus sequences in rabbit, primate and sloth genomes (Gifford *et al.*, 2008; Gilbert *et al.*, 2009; Katzourakis *et al.*, 2009), assuming that they have co-diverged, indicates that the retroviruses are tens of millions of years old. Recently, statistical methods have been used to contest the conclusion that hantaviruses and their hosts have codiverged and are of similar ages (Kang *et al.*, 2009; Ramsden *et al.*, 2008, 2009) yet others claim that valid and realistic methods to test for codivergence have not yet been developed (Schardl *et al.*, 2008).

Short-term rates are mostly estimated by statistical inference methods, which are usually described as rigorous, as indeed they are in the sense that they have been derived using explicit rules. But how certain are we that those rules are always appropriately congruent with the biological rules they seek to describe? They are obviously appropriate for estimating the short-term rate of evolutionary change of populations, which was amusingly described by Campbell (1993) as 'nothing but a soap opera. Its actors...forever changing and adapting to crisis after crisis but never getting anywhere'. The estimates produced by these methods agree with serial sampling studies, and are usually biologically sensible and supported by other evidence, much of which is recent and reliable. So, should we therefore conclude, as some have (Harkins *et al.*, 2009; Holmes, 2009), that because the short- and long-term rates are so different, we must choose between them and discard the long-term rates as probably wrong? A corollary of this conclusion would be that we also accept that most single-stranded (ss)RNA viruses are modern and originated in the

past few centuries or millennia at most. It is true that the long-term rates are mostly based on circumstantial evidence using logic similar to that used in classical studies of evolution; however, those studies turned out to be surprisingly robust. The other alternative is that there is no dilemma, and both short- and long-term rate estimates are correct and reflect different facets of the truth. The short-term rates (i.e. microevolution) reflect changes that dominate contemporary population evolution and perhaps cannot be reliably extrapolated at present into deep evolutionary time (i.e. macroevolution), in which only changes that permit a virus to adapt to new hosts, vectors, environments or gene combinations have persisted, and those that dominate population evolutionary change have become saturated or lost. We therefore discuss below the factors that may reconcile the difference in rates.

Short-term rates – microevolution

The maximum rate at which an organism can evolve is determined by the rate that its genes mutate, whereas its actual evolutionary rate depends on what proportion of the mutants survive various selective processes and appear in subsequent generations. If selection is 'neutral' and all mutants survive, then the evolutionary rate of the population is the same as its mutation rate, whereas if none do, then the evolutionary rate is zero. Of course, in practice, evolutionary rates lie somewhere between these two extremes of overwhelming positive or negative selection. Thus, the reason that estimates of the short-term evolutionary rates of many viruses are so similar to one another is probably because they are close to a maximum rate set by the biochemistry of mutation. The average mutation rates of ssRNA genes have been found to be mostly within one order of magnitude of 10^{-4} ns per site per replication. The best direct estimates obtained for plant viruses (2.4–3.0×10^{-5}) (Malpica *et al.*, 2002; Sanjuán *et al.*, 2009) fall in the lower end of this range, leading the authors to speculate that differences in selective pressures may have driven plant virus mutation rates to values lower than those for bacterial or animal ssRNA viruses. We can assume that there are around ten replication rounds per year. Note that we do not refer here to the number of replication rounds of the genome in the infected cell but, rather, to the evolutionarily significant number of generations per year over different hosts during an epidemic episode; this depends primarily on the number of inter-host transmission events. The apparent infection rates in disease progress curves (r) indicate the number of new infections per infected host in a unit of time. For a range of aphid-transmitted viruses, epidemio-logical evidence is that the mean $r \approx 0.05$ (Alonso-Prados *et al.*, 2003). This means that in a 3–4 month season, there are 4.5–6 generations. It is likely that transmission between hosts cannot occur without full infection of a source leaf, which involves fewer than 10 infection cycles (González-Jara *et al.*, 2009). This led to our conclusion that the number of infection rounds per year will be in the order of

tens and that, in nature, we can expect a maximum evolutionary rate of around 10^{-3} ns per site per year. So the new reports of the short-term evolutionary rates of plant viruses (Table 1) are close to the maximum, although data on population diversity of different ssRNA plant viruses suggest that they are somewhat more genetically stable than their animal counterparts (García-Arenal et al., 2001, 2003).

Incidentally, one of the surprises of the evolutionary rate studies has been the finding that the short-term rate estimates for viruses with ssDNA genomes are closely similar to those with ssRNA genomes. It is thought that most substitutions in ssRNA genomes result from errors introduced by the RNA-dependent RNA polymerases (RdRps) by which they replicate, as RdRps have no proof-reading mechanisms to correct for transcriptional errors. In contrast, the double-stranded (ds)DNA genomes of cellular organisms and some DNA phages mutate at around one millionth of the rate of ssRNA genomes. This slower rate reflects the fact that DNA-dependent DNA polymerases have proof-reading mechanisms which correct transcriptional errors (Domingo & Holland, 1997). However, the ssRNA-like mutation rates reported for ssDNA plant viruses are unexpected because ssDNA genomes are reported to use host cellular DNA replication enzymes for their replication (Gutierrez, 1999), and would therefore be expected to have the much slower evolutionary rates of dsDNA genes. The short-term evolutionary rates of ssDNA plant viruses (Arguello-Astorga et al., 2007) are similar to those reported for other viruses with ssDNA genomes including the bacteriophage ΦX174 (Raney et al., 2004) and several anello-, circo- and parvoviruses of animals (Duffy et al., 2008; van der Walt et al., 2008). This suggests that either the proof-reading ability of replicases may not be the only factor influencing the underlying mutation rate (Duffy et al., 2008) or our understanding of the replication of ssDNA genomes is incomplete.

Long-term rates – macroevolution

What factors might explain the difference between short- and long-term rates? Is the rate timescale-dependent in the way that the genes of large cellular organisms have been found to be? An apparent deceleration of evolutionary rates over increasing timescales has been reported for several organisms, and the interpretation of this phenomenon is an area of very active research at present (Debruyne & Poinar, 2009; Ho et al., 2005, 2007). The difference in rates found with viruses is much larger than those reported for cellular organisms, but this may merely reflect the difference between ssRNA and dsDNA genomes.

Many factors might produce a timescale dependence of evolutionary rates by severely restricting protein changes or combinations of changes, to those that maintain function, yet provide sufficient novelty for a virus to track its ecological niche and exploit new opportunities as they arise. The following paragraphs provide examples of these.

One third of the primary mutants in ssRNA genomes are multiple and about two thirds of them are insertions and deletions (Malpica et al., 2002); most are likely to be very deleterious or immediately lethal. A large fraction (~70 %) of point substitutions are deleterious or lethal (Carrasco et al., 2007; Sanjuan et al., 2004) meaning that most mutations in RNA genomes are not likely to survive, particularly the non-synonymous changes (Hughes, 2009). Similar effects have been reported for animal viruses (Pybus et al., 2007).

Only a small number of sites in a gene vary, and most of these will probably be quickly saturated. Wu et al. (2008) noted that in the study of TYLCV from China (Ge et al., 2007) only 'four nucleotide positions accounted for close to half (18 of 41) of the observed substitutions'. Furthermore, the sites varying in one species may differ from those varying in another. We have analysed the published CP sequences of tobamoviruses (Gibbs et al., 2008a) to find the mutually informative pattern of amino acids (Chiu & Kolodziejczak, 1991) that distinguish TMV from tomato mosaic tobamovirus (ToMV), which is its sister species. Of the amino acids in 67 different TMV CPs, 82 % are invariant or differ in only a single sequence, only five of 159 codons (3 %) define the species, and only one of them was among the five that define the cluster of ToMV CPs. Similar effects have been reported for influenza virus, where the sites that vary change both during the evolution of a single lineage (Smith et al., 2004) and when the virus infects different host species (Gibbs et al., 2007). A clear example of the fact that viral functions, even those critical for the survival of a virus, do not necessarily evolve in a linear manner using a single set of codons is shown by the wanderings of the 'Caspar carboxylates' which bind the CPs in the virions of each tobamovirus species, yet relax correctly to allow infection to occur (Stubbs, 1999).

Current modelling methods seem to be better suited to analysing evolution resulting from phyletic gradualism rather than punctuated equilibria (Eldredge & Gould, 1972), yet evolutionary rates probably vary greatly within viral lineages as selection modes and intensities vary (Thorne & Kishino, 2005). There is much evidence from animal virus studies that evolutionary rates increase when viruses change host (Cao et al., 1995; Graff et al., 1994; Itoh et al., 1997; Sawyer et al., 1994), and the phylogenetic consequences of this have been well studied in influenza hemagglutinin proteins that have adapted to grow in embryonated hen's eggs (Bush, 2004; Bush et al., 1999, 2000). Methods for specifically detecting punctuations, otherwise called heterotachy, have been developed (Lopez et al., 2002; Pagel & Meade, 2008) but there are as yet few reports of their application to viral sequences (Dorman, 2007).

The topology of the phylogenies of many RNA viruses indicates that the birth and death of lineages is a dominant feature of their evolution (Holmes, 2009). Populations probably quickly lose ancient sister lineages that would

enable the deep divergences to be detected phylogenetically. The death of lineages may result from random processes, but also, for some viruses, from active competition for healthy susceptible hosts and the effects of cross protection or super infection immunity (McKinney, 1929; Salaman, 1933; Thung, 1931).

The long-term effective population sizes of viruses are probably small, probably as a result of severe bottlenecking during transmission between hosts (Betancourt et al., 2008; French & Stenger, 2003; Moury et al., 2007; Sacristán et al., 2003). A recent analysis (Hughes, 2009) of the population genetics of a large sample of complete potyvirus genomes has found that their effective populations are in the order of 10^4. This, and the birth–death topology, suggest that despite large populations within individual infected plants, only small numbers survive the hazards of being transmitted to new hosts; there is probably strong selection for survival of the ecologically fittest.

Protein evolution

Methods for sequencing proteins were developed before those for sequencing nucleic acids, and the initial concept of molecular clocks was developed using those sequences (Hartl & Dykhuizen, 1979; Sarich & Wilson, 1967; Wilson et al., 1977; Zuckerkandl & Pauling, 1962, 1965). Proteins populate most of the interface between viruses and their hosts and vectors. Virus proteins will therefore be constrained to evolve in synchrony with those of their hosts and vectors, whose proteins are encoded in DNA; therefore, they will evolve slowly. This may be the reason why the amino acid motifs of viruses are phylogenetically persistent. There is no reason to suspect that proteins encoded by RNA genomes evolve in an inherently different manner or at an inherently different rate from those encoded by DNA genomes. Furthermore, although the mutation rates of nucleic acids cover at least a million-fold range, those of individual proteins and protein families cover less than a thousand-fold range (Luz & Vingron, 2006; Wilson et al., 1977; Wolf et al., 2008). So, if we assume that the rates of evolution of viral proteins fall within the limits estimated for all proteins, then we can obtain a broad estimate of the possible age for each viral protein family. Three decades ago, using the sequence differences of seven tobamovirus CPs and the estimated evolutionary rates of 50 protein families (Wilson et al., 1977), this approach suggested that the tobamoviruses first diverged around the same time as the flowering plants (Gibbs, 1980). An increased knowledge of tobamovirus CP sequences and protein substitution rates has had little effect on that estimate.

Sequence and structural comparisons of proteins are revealing the deep phylogenetic relationships of viruses (Koonin & Dolja, 2006; Koonin et al., 2008) but methods for dating those phylogenies are only just emerging.

Why bother with rates and dates?

There are important practical reasons for understanding the rates of evolution of different plant viruses. In recent times, there have been massive global changes in the ecology of plants, their viruses and vectors. Many of these changes have resulted from the invention and spread of agriculture over the past few millennia and post-Columban world trade over the past 500 years. Dating the recent prehistory of individual viruses during this maelstrom may provide important new insights into understanding their ecology and control. For example, it may enable a cost/benefit analysis of quarantine measures to be made. A study of Australian potyviruses showed that around half of these, found only in crops, arrived in Australia from overseas since Europeans colonized the continent 200 years ago (Gibbs et al., 2008b, c, d). In contrast, only two or three potyviruses have spread to crops from endemic Australian plants infected with an older immigrant potyvirus population over the same period (Webster et al., 2007). Thus, the recent immigrants are by far the major source of potyviruses for Australian crops and, on average, one new potyvirus evades quarantine and becomes established in Australian crops every decade, which is about the same rate that new potyviruses enter the UK (Jones & Baker, 2007). These studies also indicate that more potyviruses may be seed-borne than had been previously reported. If past quarantine practices are continued, this rate of quarantine evasion is likely to continue as there is a large pool of potyviruses found elsewhere in the world that are not yet recorded in Australia. Similar studies of the viruses of the tallgrass prairie reserve of north-eastern Oklahoma (Muthukumar et al., 2009) and viruses of Costa Rica (Wren et al., 2006) are being undertaken.

Knowledge of when a crop species first encountered a particular virus also provides a new dimension to the study of its resistance genes and may reveal the mode and rate that such genes are generated.

Knowledge of dates and rates may also be of practical value to those attempting to enhance agricultural production using molecular techniques. For example, the ability of transgenes to protect plants against particular viral infections is primarily related to the homology of the genes involved (Tripathi et al., 2008); this is, of course, related to their date of divergence, so a knowledge of the rate of evolutionary change of viral genes will indicate how long a particular transgene might remain effective. Finally, sequences lodged in the international databases provide a progressive and dated set of samples of the world virus populations. Analysis of those sequences, their dates and their diversity, can aid the design of genus-specific primers (Zheng et al., 2008) and will soon enable estimates of the total world population of particular viral taxa, known and unknown, to be made.

Acknowledgements

References

Almeida, R. P. 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.

Alonso-Prados, J. L., Luis-Arteaga, M., Alvarez, J. M., Moriones, E., Batlle, A., Laviña, A., García-Arenal, F. & Fraile, A. (2003). Epidemics of aphid-transmitted viruses in melon crops in Spain. *Eur J Plant Pathol* **109**, 129–138.

Arguello-Astorga, G., Ascencio-Ibáñez, J. T., Dallas, M. B., Orozco, B. M. & Hanley-Bowdoin, L. (2007). High-frequency reversion of geminivirus replication protein mutants during infection. *J Virol* **81**, 11005–11015.

Ashby, M. K., Warry, A., Bejarano, E. R., Khashoggi, A., Burrell, M. & Lichtenstein, C. (1997). Analysis of multiple copies of geminiviral DNA in the genomes of four closely related Nicotiana species suggest a unique integration event. *Plant Mol Biol* **35**, 313–321.

Baker, P. F. (1960). Aphid behaviour on healthy and on yellows-virus-infected sugar beet. *Ann Appl Biol* **48**, 384–391.

Bejarano, E. R., Khashoggi, A., Witty, M. & Lichtenstein, C. (1996). Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. *Proc Natl Acad Sci U S A* **93**, 759–764.

Bellwood, P. (2005). *First Farmers: the Origins of Agricultural Societies*. Malden, MA: Blackwell Publishing.

Betancourt, M., Fereres, A., Fraile, A. & García-Arenal, F. (2008). Estimation of the effective number of founders that initiate an infection after aphid transmission of a multipartite plant virus. *J Virol* **82**, 12416–12421.

Blok, J., Mackenzie, A., Guy, P. & Gibbs, A. (1987). Nucleotide sequence comparisons of turnip yellow mosaic isolates from Australia and Europe. *Arch Virol* **97**, 283–295.

Burroughs, W. J. (2005). *Climate Change in Prehistory: the End of the Reign of Chaos*. Cambridge: Cambridge University Press.

Bush, R. M. (2004). Influenza as a model system for studying the cross-species transfer and evolution of the SARS coronavirus. *Philos Trans R Soc Lond B Biol Sci* **359**, 1067–1073.

Bush, R. M., Fitch, W. M., Bender, C. A. & Cox, N. J. (1999). Positive selection on the H3 hemagglutinin gene of human influenza virus A. *Mol Biol Evol* **16**, 1457–1465.

Bush, R. M., Smith, C. B., Cox, N. J. & Fitch, W. M. (2000). Effects of passage history and sampling bias on phylogenetic reconstruction of human influenza A evolution. *Proc Natl Acad Sci U S A* **97**, 6974–6980.

Campbell, J. H. (1993). A tilt at cladism or let's contemplate evolution instead of our belly buttons. *Mem Ass Aust Palaeontol* **15**, 43–50.

Cao, J. X., Ni, H., Wills, M. R., Campbell, G. A., Sil, B. K., Ryman, K. D., Kitchen, I. & Barrett, A. D. (1995). Passage of Japanese encephalitis virus in HeLa cells results in attenuation of virulence in mice. *J Gen Virol* **76**, 2757–2764.

Carrasco, P., de la Iglesia, F. & Elena, S. F. (2007). The distribution of fitness and virulence effects caused by single nucleotide substitutions in tobacco etch virus. *J Virol* **81**, 12979–12984.

Chiu, D. K. & Kolodziejczak, T. (1991). Inferring consensus structure from nucleic acid sequences. *Comput Appl Biosci* **7**, 347–352.

Clarkson, J. J., Knapp, S., García, V. F., Olmstead, R. G., Leitch, A. R. & Chase, M. W. (2004). Phylogenetic relationships in Nicotiana (Solanaceae) inferred from multiple plastid DNA regions. *Mol Phylogenet Evol* **33**, 75–90.

Clarkson, J. J., Lim, K. Y., Kovarik, A., Chase, M. W., Knapp, S. & Leitch, A. R. (2005). Long-term genome diploidization in allopolyploid Nicotiana section Repandae (Solanaceae). *New Phytol* **168**, 241–252.

Crosby, A. W. (2004). *Ecological Imperialism: the Biological Expansion of Europe, 900–1900*. Cambridge: Cambridge University Press.

Debruyne, R. & Poinar, H. N. (2009). Time dependency of molecular rates in ancient DNA data sets, a sampling artifact? *Syst Biol* **58**, 348–360.

Domingo, E. & Holland, J. J. (1997). RNA virus mutations and fitness for survival. *Annu Rev Microbiol* **51**, 151–178.

Dorman, K. S. (2007). Identifying dramatic selection shifts in phylogenetic trees. *BMC Evol Biol* **7**, S10.

Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics* **148**, 1667–1686.

Drummond, A. J., Pybus, O. G., Rambaut, A., Forsberg, R. & Rodrigo, A. G. (2003). Measurably evolving populations. *Trends Ecol Evol* **18**, 481–488.

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.

Duffy, S. & Holmes, E. C. (2009). Validation of high rates of nucleotide substitution in geminiviruses: phylogenetic evidence from East African cassava mosaic viruses. *J Gen Virol* **90**, 1539–1547.

Duffy, S., Shackelton, L. A. & Holmes, E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* **9**, 267–276.

Eldredge, N. & Gould, S. J. (1972). Punctuated equilibria: an alternative to phyletic gradualism. In *Models in Paleobiology*, pp. 82–115. Edited by T. J. M. Schopf. San Fransisco: Freeman Cooper.

Fargette, D., Pinel-Galzi, A., Sérémé, D., Lacombe, S., Hébrard, H., Traoré, O. & Konaté, G. (2008a). Diversification of rice yellow mottle virus and related viruses spans the history of agriculture from the Neolithic to the present. *PLoS Pathog* **4**, e1000125.

Fargette, D., Pinel, A., Rakotomalala, M., Sangu, E., Traoré, O., Sérémé, D., Sorho, F., Issaka, S., Hébrard, E. & other authors (2008b). Rice yellow mottle virus, an RNA plant virus, evolves as rapidly as most RNA animal viruses. *J Virol* **82**, 3584–3589.

Fourment, M. & Gibbs, M. J. (2006). PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. *BMC Evol Biol* **6**, 1.

Fourment, M., Gibbs, A. J. & Gibbs, M. J. (2008). SWEBLAST: a sliding window web-based BLAST tool for recombinant analysis. *J Virol Methods* **152**, 98–101.

Fraile, A., Malpica, J. M., Aranda, M. A., Rodríguez-Cerezo, E. & García-Arenal, F. (1996). Genetic diversity in tobacco mild green mosaic tobamovirus infecting the wild plant Nicotiana glauca. *Virology* **223**, 148–155.

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.

French, R. & Stenger, D. C. (2003). Evolution of wheat streak mosaic virus: dynamics of population growth within plants may explain limited variation. *Annu Rev Phytopathol* **41**, 199–214.

García-Arenal, F., Fraile, A. & Malpica, J. M. (2001). Variability and genetic structure of plant virus populations. *Annu Rev Phytopathol* **39**, 157–186.

García-Arenal, F., Fraile, A. & Malpica, J. M. (2003). Variation and evolution of plant virus populations. *Int Microbiol* **6**, 225–232.

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. J. (1980). How ancient are the tobamoviruses? *Intervirology* **14**, 101–108.

Gibbs, A. J., Blok, J., Coates, D. J., Guy, P. L., Mackenzie, A. & Pigram, N. (1986). Turnip yellow mosaic virus in an endemic Australian alpine Cardamine. In *Flora and Fauna of Alpine Australasia; Ages and Origins*, pp. 289–300. Edited by B. A. Barlow. Collingwood, Australia: CSIRO.

Gibbs, A., Blok, J., Guy, P., Keese, P., Mackenzie, A. & Meek, D. (1989). Turnip yellow mosaic virus – a significant migrant to the Australian Alps. In *The Scientific Significance of the Australian Alps*, pp. 385–386. Edited by R. Good. Canberra: Australian Academy of Science.

Gibbs, M. J., Wayper, P., Fourment, M. L., Wood, J. T., Ohshima, K., Armstrong, J. S. & Gibbs, A. J. (2007). The variable codons of H3 influenza A virus haemagglutinin genes. *Arch Virol* 152, 11–24.

Gibbs, A. J., Gibbs, M. J., Ohshima, K. & García-Arenal, F. (2008a). More plant virus evolution; past present and future. In *Origin and Evolution of Viruses*, 2nd edn, pp. 229–250. Edited by E. Domingo, C. R. Parrish & J. J. Holland. London: Academic Press.

Gibbs, A. J., Mackenzie, A. M., Wei, K.-J. & Gibbs, M. J. (2008b). The potyviruses of Australia. *Arch Virol* 153, 1411–1420.

Gibbs, A. J., Ohshima, K., Phillips, M. J. & Gibbs, M. J. (2008c). The prehistory of potyviruses: their initial radiation was during the dawn of agriculture. *PLoS One* 3, e2523.

Gibbs, A. J., Trueman, J. W. H. & Gibbs, M. J. (2008d). The bean common mosaic virus lineage of potyviruses: where did it arise and when? *Arch Virol* 153, 2177–2188.

Gifford, R. J., Katzourakis, A., Tristem, M., Pybus, O. G., Winters, M. & Shafer, R. W. (2008). A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. *Proc Natl Acad Sci U S A* 105, 20362–20367.

Gilbert, C., Maxfield, D. G., Goodman, S. M. & Feschotte, C. (2009). Parallel germline infiltration of a lentivirus in two Malagasy lemurs. *PLoS Genetics* 5, e1000425.

González-Jara, P., Fraile, A., Canto, T. & García-Arenal, F. (2009). The multiplicity of infection of a plant virus varies during colonisation of its eukaryotic host. *J Virol* 83, 7487–7494.

Graff, J., Kasang, C., Normann, A., Pfisterer-Hunt, M., Feinstone, S. M. & Flehmig, B. (1994). Mutational events in consecutive passages of hepatitis A virus strain GBM during cell culture adaptation. *Virology* 204, 60–68.

Guindon, S. & Gascuel, O. (2003). PhyML – a simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52, 696–704.

Gutierrez, C. (1999). Geminivirus DNA replication. *Cell Mol Life Sci* 56, 313–329.

Guy, P. & Gibbs, A. J. (1981). A tymovirus of Cardamine sp. from alpine Australia. *Australas Plant Pathol* 10, 12–13.

Hanada, K., Suzuki, Y. & Gojobori, T. (2004). A large variation in the rates of synonymous substitution for RNA viruses and its relationships to a diversity of viral infection and transmission modes. *Mol Biol Evol* 21, 1074–1080.

Harkins, G. W., Delport, W., Duffy, S., Wood, N., Monjane, A. L., Owor, B. E., Donaldson, L., Saumtally, S., Triton, G. & other authors (2009). Experimental evidence indicating that mastreviruses probably did not co-diverge with their hosts. *Virol J* 6, 104.

Harlan, J. R. (1976). Diseases as a factor in plant evolution. *Annu Rev Phytopathol* 14, 31–51.

Hartl, D. & Dykhuizen, D. (1979). A selectively driven molecular clock. *Nature* 281, 230–231.

Hayden, C. M., Mackenzie, A. M. & Gibbs, A. J. (1998). Virion protein sequence variation among Australian isolates of turnip yellow mosaic virus. *Arch Virol* 143, 191–201.

Ho, S. Y. W., Phillips, M. J., Cooper, A. & Drummond, A. J. (2005). Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol Biol Evol* 22, 1561–1568.

Ho, S. Y. W., Shapiro, B., Phillips, M. J., Cooper, A. & Drummond, A. J. (2007). Evidence for time dependency of molecular rate estimates. *Syst Biol* 56, 515–522.

Holmes, F. O. (1951). Indications of a New World origin of tobacco mosaic virus. *Phytopathology* 41, 341–349.

Holmes, E. C. (2009). *The Evolution and Emergence of RNA viruses*. Oxford: Oxford University Press.

Hughes, A. L. (2009). Small effective population sizes and rare nonsynonymous variants in potyviruses. *Virology* 393, 127–134.

Itoh, M., Isegawa, Y., Hotta, H. & Homma, M. (1997). Isolation of an avirulent mutant of Sendai virus with two amino acid mutations from a highly virulent field strain through adaptation to LLC-MK2 cells. *J Gen Virol* 78, 3207–3215.

Jenkins, G. M., Rambaut, A., Pybus, O. G. & Holmes, E. C. (2002). Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J Mol Evol* 54, 156–165.

Jiu, M., Zhou, X.-P., Tong, L., Xu, J., Yang, X., Wan, F.-H. & Liu, S.-S. (2007). Vector-virus mutualism accelerates population increase of an invasive whitefly. *PLoS One* 2, e182.

Jones, R. A. C. (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.

Jones, D. R. & Baker, R. H. A. (2007). Introductions of non-native plant pathogens into Great Britain, 1970–2004. *Plant Pathol* 56, 891–910.

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.

Katzourakis, A., Gifford, R. J., Tristem, M., Thomas, M., Gilbert, P. & Pybus, O. G. (2009). Macroevolution of complex retroviruses. *Science* 325, 1512.

Kirino, N., Inoue, K., Tanina, K., Yamazaki, Y. & Ohki, S. T. (2008). *Turnip yellow mosaic virus* isolated from Chinese cabbage in Japan. *J Gen Plant Pathol* 74, 331–334.

Koonin, E. V. & Dolja, V. V. (2006). Evolution of complexity in the viral world: the dawn of a new vision. *Virus Res* 117, 1–4.

Koonin, E. V., Wolf, Y. I., Nagasaki, K. & Dolja, V. V. (2008). The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. *Nat Rev Microbiol* 6, 925–939.

Lopez, P., Casane, D. & Philippe, H. (2002). Heterotachy, an important process of protein evolution. *Mol Biol Evol* 19, 1–7.

Luz, H. & Vingron, M. (2006). Family specific rates of protein evolution. *Bioinformatics* 22, 1166–1171.

Malmstrom, C. M., Shu, R., Linton, E. W., Newton, L. A. & Cook, M. A. (2007). Barley yellow dwarf viruses (BYDVs) preserved in herbarium specimens illuminate historical disease ecology of invasive and native grasses. *J Ecol* 95, 1153–1166.

Malpica, J. M., Fraile, A., Moreno, I., Obies, C. I., Drake, J. W. & García-Arenal, F. (2002). The rate and character of spontaneous mutation in an RNA virus. *Genetics* 162, 1505–1511.

Martin, D. P., Williamson, C. & Posada, D. (2005). RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* 21, 260–262.

McKinney, H. H. (1929). Mosaic diseases in the Canary Islands, West Africa, and Gibraltar. *J Agric Res* 39, 557–578.

McKinney, H. H. (1937). *Mosaic Diseases of Wheat and Related Cereals*, pp. 1–23. US Department of Agriculture Circular No. 442.

Moury, B., Fabre, F. & Senoussi, R. (2007). Estimation of the number of virus particles transmitted by an insect vector. *Proc Natl Acad Sci U S A* **104**, 17891–17896.

Murad, L., Bielawski, J. P., Matyasek, R., Kovaŕík, A., Nichols, R. A., Leitch, A. R. & Lichtenstein, C. P. (2004). The origin and evolution of geminivirus-related DNA sequences in Nicotiana. *Heredity* **92**, 352–358.

Murphy, D. J. (2007). *People, Plants and Genes: the Story of Crops and Humanity*. Oxford: Oxford University Press.

Muthukumar, V., Melcher, U., Pierce, M., Wiley, G. B., Roe, B. A., Palmer, M. W., Thapa, V., Ali, A. & Ding, T. (2009). Non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma frequently contain virus-like sequences in particulate fractions. *Virus Res* **141**, 169–173.

Pagel, M. & Meade, A. (2008). Modelling heterotachy in phylogenetic inference by reversible-jump Markov chain Monte Carlo. *Philos Trans R Soc Lond B Biol Sci* **363**, 3955–3964.

Pybus, O. G., Rambaut, A., Belshaw, R., Freckleton, R. P., Drummond, A. J. & Holmes, E. C. (2007). Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. *Mol Biol Evol* **24**, 845–852.

Ramsden, C., Melo, F. L., Figueiredo, L. M., Holmes, E. C. & Zanotto, P. M. (2008). High rates of molecular evolution in hantaviruses. *Mol Biol Evol* **25**, 1488–1492.

Ramsden, C., Holmes, E. C. & Charleston, M. A. (2009). Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. *Mol Biol Evol* **26**, 143–153.

Raney, J. L., Delongchamp, R. R. & Valentine, C. R. (2004). Spontaneous mutant frequency and mutation spectrum for gene A of ΦX174 grown in *E. coli*. *Environ Mol Mutagen* **44**, 119–127.

Rodriguez-Cerezo, E. & García-Arenal, F. (1989). Genetic heterogeneity of the RNA genome population in the plant virus U5–TMV. *Virology* **170**, 418–423.

Ruddiman, W. F. (2003). The anthropogenic greenhouse era began thousands of years ago. *Clim Change* **61**, 261–293.

Rybicki, E. P. (1994). A phylogenetic and evolutionary justification for three genera of *Geminiviridae*. *Arch Virol* **139**, 49–77.

Sacristán, S., Malpica, J. M., Fraile, A. & García-Arenal, F. (2003). Estimation of population bottlenecks during systemic movement of *Tobacco mosaic virus* in tobacco plants. *J Virol* **77**, 9906–9911.

Salaman, R. N. (1933). Protective inoculation against a plant virus. *Nature* **131**, 468.

Sanjuan, R., Moya, A. & Elena, S. F. (2004). The contribution of epistasis to the architecture of fitness in an RNA virus. *Proc Natl Acad Sci U S A* **101**, 15376–15379.

Sanjuán, R., Agudelo-Romero, P. & Elena, S. F. (2009). Upper-limit mutation rate estimation for a plant RNA virus. *Biol Lett* **5**, 394–396.

Sarich, V. M. & Wilson, A. C. (1967). Rates of albumin evolution in primates. *Proc Natl Acad Sci U S A* **58**, 142–148.

Sawyer, L. S., Wrin, M. T., Crawford-Miksza, L., Potts, B., Wu, Y., Weber, P. A., Alfonso, R. D. & Hanson, C. V. (1994). Neutralization sensitivity of human immunodeficiency virus type 1 is determined in part by the cell in which the virus is propagated. *J Virol* **68**, 1342–1349.

Schardl, C. L., Craven, D., Speakman, S., Stromberg, A., Lindstrom, A. & Yoshida, R. (2008). A novel test for host-symbiont codivergence indicates ancient origin of fungal endophytes in grasses. *Syst Biol* **57**, 483–498.

Simmons, H. E., Holmes, E. C. & Stephenson, A. G. (2008). Rapid evolutionary dynamics of zucchini yellow mosaic virus. *J Gen Virol* **89**, 1081–1085.

Smith, D. J., Lapedes, A. S., de Jong, J. C., Bestebroer, T. M., Rimmelzwaan, G. F., Osterhaus, A. D. & Fouchier, R. A. (2004). Mapping the antigenic and genetic evolution of influenza virus. *Science* **305**, 371–376.

Soltis, D. E., Bell, C. D., Kim, S. & Soltis, P. S. (2008). Origin and early evolution of angiosperms. *Ann N Y Acad Sci* **1133**, 3–25.

Stenger, D. C., Seifers, D. L. & French, R. (2002). Patterns of polymorphism in wheat streak mosaic virus: sequence space explored by a clade of closely related viral genotypes rivals that between the most divergent strains. *Virology* **302**, 58–70.

Stubbs, G. (1999). Tobacco mosaic virus particle structure and the initiation of disassembly. *Philos Trans R Soc Lond B Biol Sci* **354**, 551–557.

Stukenbrock, E. H. & McDonald, B. A. (2008). The origins of plant pathogens in agro-ecosystems. *Annu Rev Phytopathol* **46**, 75–100.

Sweeney, M. & McCouch, S. (2007). The complex history of the domestication of rice. *Ann Bot* **100**, 951–957.

Thorne, J. L. & Kishino, H. (2005). *Estimation of Divergence Times from Molecular Sequence Data.* Vienna: Springer.

Thung, T. H. (1931). Smetstof en plantencel bij enkele virusziekten van de tabaksplant. *Eur J Plant Pathol* **45**, 247–259 (in Dutch).

Tripathi, S., Suzuki, J. Y., Ferreira, S. & Gonsalves, D. (2008). Papaya ringspot virus-P: characteristics, pathogenicity, sequence variability and control. *Mol Plant Pathol* **9**, 269–280 (in Dutch).

van der Walt, E., Martin, D. P., Varsani, A., Polston, J. E. & Rybicki, E. P. (2008). Experimental observations of rapid maize streak virus evolution reveal a strand-specific nucleotide substitution bias. *Virol J* **5**, 104.

Vavilov, N. I. (1940). The theory of origins of cultivated plants after Darwin. In *Origin and Geography of Cultivated Plants*, p. 498. Cambridge: Cambridge University Press.

Webster, C. G., Coutts, B. A., Jones, R. A. C., Jones, M. G. K. & Wylie, S. J. (2007). Virus impact at the ancient ecosystem-recent agroecosystem interface: studies on three legume-infecting potyviruses in the Southwest Australian Floristic Region. *Plant Pathol* **56**, 729–742.

Wertheim, J. O. & Worobey, M. (2009). Dating the age of the SIV lineages that gave rise to HIV-1 and HIV-2. *PLoS Comput Biol* **5**, e1000377.

Wilson, A. C., Carlson, S. S. & White, T. J. (1977). Biochemical evolution. *Annu Rev Biochem* **46**, 573–639.

Wolf, M. Y., Wolf, Y. I. & Koonin, E. V. (2008). Comparable contributions of structural–functional constraints and expression level to the rate of protein sequence evolution. *Biol Direct* **3**, 40.

Wolfe, N. D., Dunavan, C. P. & Diamond, J. (2007). Origins of major human infectious diseases. *Nature* **447**, 279–283.

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., Melcher, U., Guo, X., Wang, X., Fan, L. & Zhou, G. (2008). Assessment of codivergence of *Mastreviruses* with their plant hosts. *BMC Evol Biol* **8**, 335.

Zheng, L., Wayper, P. J., Gibbs, A. J., Fourment, M., Rodoni, B. C. & Gibbs, M. J. (2008). Accumulating variation at conserved sites in potyvirus genomes is driven by species discovery and affects degenerate primer design. *PLoS One* **3**, e1586.

Zuckerkandl, E. & Pauling, L. B. (1962). Molecular disease, evolution, and genetic heterogeneity. In *Horizons in Biochemistry*, pp. 189–225. Edited by M. Kasha & B. Pullman. New York: Academic Press.

Zuckerkandl, E. & Pauling, L. B. (1965). Evolutionary divergence and convergence in proteins. In *Evolving Genes and Proteins*, pp. 97–166. Edited by V. Bryson & H. J. Vogel. New York: Academic Press.