THE ROYAL SOCIETY

Evolution and origins of tobamoviruses

Adrian Gibbs

Research School of Biological Sciences, Australian National University, PO Box 475, ACT 2601, Australia (gibbs@rsbs.anu.edu.au)

More than a dozen tobamoviruses are known. In nature, each species probably survives by moving between several closely related host species. Each infected plant contains a population of variants, but in most host populations the tobamovirus population is stable. The phylogenetic relationships of tobamovirus species broadly correlate with those of their angiosperm hosts. The simplest explanation for this correlation is that they have coevolved with the angiosperms, and hence, like them, are about 120–140 million years old. Gene sequence differences between species also indicate that the tobamoviruses are an ancient genus. Their gene sequences, and the protein motifs they encode, link them to tobraviruses, hordeiviruses and soil-borne wheat mosaic virus, more distantly to the tricornaviruses, and even to hepatitis virus E and other furoviruses, rubiviruses and alphaviruses. Their progenitors may have been associated with charophycean algae, and perhaps also plasmodiophoromycete fungi.

Keywords: tobamoviruses; coevolution; alpha-like viruses

1. INTRODUCTION

Evolution is the process by which the genetic structure of the population of an organism changes with time. The pattern and speed of evolution is inferred most frequently by comparing distinct lineages of organisms that have been produced by past evolutionary changes, but it can also be observed directly in species that evolve fast, such as some viruses, bacteria and domesticated organisms. The tobamoviruses have been studied in both ways and the evolutionary changes assessed, especially recently, by gene sequencing. A consistent story of tobamovirus origins and evolution is emerging.

2. EVOLUTION IN INOCULATED PLANTS

Early studies (see pp. 292–311 of Bawden (1964)) demonstrated that viruses were genetically variable. It was found that some characters of tobacco mosaic tobamovirus (TMV) isolates may be altered by culturing them at different temperatures (Holmes 1934; Kassanis 1957; Jones & Dawson 1978) or in different plant species (Johnson 1947; Yarwood 1979). The timing of the genetic changes resulting in these differences was uncertain as the experiments did not distinguish whether they had occurred during the experiments or resulted from selection among variants pre-existing in the inocula. However Aldaoud *et al*. (1989) removed some of the uncertainty by using a TMV population obtained by transcribing a cloned DNA encoding the entire genome of the virus into infectious RNA (Dawson *et al*. 1986). The resulting inoculum contained both temperature sensitive (*ts*) and necrotic lesion (*nl*) variants that had emerged since cloning, and these were assayed independently. It was found that, although the proportion of *ts* variants in the population was not affected by growing this population at different temperatures or in different plant species, the proportion of *nl* variants was influenced significantly. The latter was decreased greatly at elevated temperatures and in some hosts, notably *Physalis floridana*, and there were also 'surprisingly large, apparently random changes in the proportion of *nl* variants . . . in individual plants.' Thus, mutation and founder effects can affect the TMV population in an individual host plant as rapidly as they affect populations of the fastest evolving animal viruses, such as equine infectious anaemia and human immunodeficiency lentiviruses (Hahn *et al*. 1986).

3. EVOLUTION IN NATURALLY INFECTED HOST POPULATIONS

F. Garcia-Arenal and his colleagues have made a series of important evolutionary comparisons of natural tobamovirus populations. In one study, the genomes of 26 isolates of pepper mild mottle tobamovirus (PMMV) were obtained from different European pepper crops over a seven-year period, and were compared by oligonucleotide fragment mapping (Rodriguez-Cerezo *et al*. 1989). The crops were found to be infected by a seemingly stable PMMV population, that persisted from year to year, and whose sequences differed, at most, by about 4%. They also studied tobacco mild green mosaic tobamovirus (TMGMV) in populations of *Nicotiana glauca*, a shrub that has migrated over the past two centuries from central South America to most regions of the world that have a Mediterranean climate (Fraile *et al*. 1996; Moya *et al*. 1993; Rodriguez-Cerezo *et al*. 1991). They found that the TMGMV populations were also closely similar, although Californian and Cretan populations were twice as variable as those in Australia and Spain. The Californian and Cretan populations were not separated in a cluster analysis, implying that they were from an older single population. By contrast, the Australian and Spanish populations formed distinct tight subclusters, and thus probably were more recently established from the older population. However the total diversity of all the populations sampled

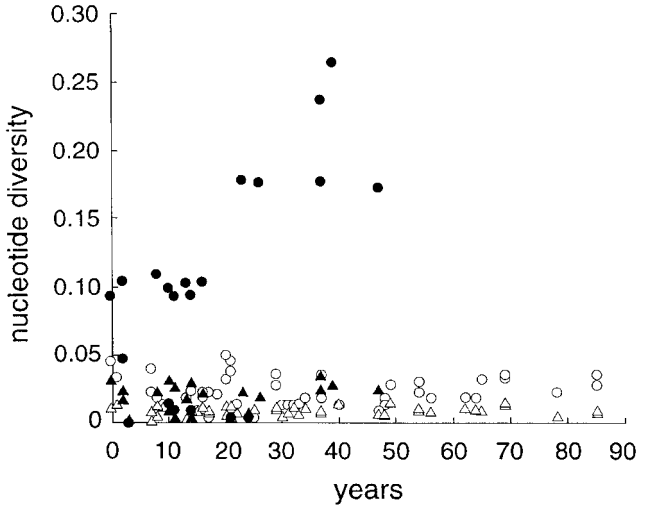

Figure 1. Relationship between nucleotide differences and time of isolation for nucleotide sequences from TMV and TMGMV isolates from herbarium specimens collected in New South Wales, Australia. The difference values for synonymous (circles) and non-synonymous (triangles) sites were calculated by the PBL method (Li 1993) for each pair of TMV (filled) or TMGMV (outlined) isolates. Redrawn from Fraile *et al.* (1997).

was no greater than that of the two most diverse, suggesting that 'There is an upper threshold for TMGMV diversity, and this may have been attained in some of these geographically defined populations' (Fraile *et al.* 1996). Most recently Garcia-Arenal's group have compared gene sequences of TMGMV isolates obtained from herbarium specimens of *N. glauca* collected in eastern Australia over the past century (Fraile *et al.* 1997) and found that the genetic diversity of TMGMV isolates did not increase over that period. These results indicate that the Australian TMGMV population is genetically stable, and its present small diversity resulted from a lack of diversity in the founder population.

Both TMV and TMGMV were isolated from the Australian *N. glauca* herbarium specimens (Fraile *et al.* 1997) collected before 1950; three of the pre-1950 samples yielded TMV alone, one TMGMV alone, four a mixture, and one a recombinant. Only TMGMV has been isolated from more recent specimens, and TMV seems to have disappeared, as it has during the same period from the *N. glauca* population of the Canary Islands; it was found there in the 1920s and 1930s (McKinney 1935) but not recently (Fraile *et al.* 1997). Gene sequences from the Australian isolates showed surprisingly that although the TMGMV population did not increase in genetic diversity over the century sampled, the TMV population did. There was a considerable increase in nucleotide changes that did not alter the encoded amino acids (i.e. synonymous changes), rather than in those that changed the encoded amino acids (i.e. non-synonymous changes) (figure 1). This may be linked to the interaction between the viruses as Fraile *et al.* (1997) showed that although TMGMV virions accumulate to the same concentration in both doubly and singly infected plants, TMV virions attain only one-tenth the concentration in doubly infected plants than in singly infected plants. Thus, a likely cause for the disappearance of TMV from the Australian *N. glauca* populations is that a 'mutational meltdown' occurred as a result of a process named Muller's rachet (Lynch *et al.* 1993). Individual TMV inocula involved in the spread of the virus from infected to healthy *N. glauca* plants must contain more than a threshhold number of genomes to ensure that each will contain at least one of the fittest genomes. Thus, when the proportion of doubly infected plants in the population exceeds a critical value, and the virion concentration of TMV in doubly infected plants falls, the TMV population will be progressively dominated by less fit genomes, and will succumb.

One might conclude from the interaction of TMV and TMGMV populations in *N. glauca* that it is the natural host of TMGMV but not of TMV, but this is probably incorrect. *N. glauca* is believed to be a native of central South America and to have been spread worldwide from there over the past two centuries, especially during the mid-19th century. However TMGMV was not isolated from any of 50 *N. glauca* samples from Argentina and Bolivia (Fraile *et al.* 1996), although in present European, North American and Australian *N. glauca* populations at least one-quarter of the plants are infected.

Similarly *N. tabacum*, which is an amphidiploid species found only in crops or as a 'crop fugitive', is unlikely to have been the original preferred host of TMV. Holmes (1950) noted that TMV infects a very wide range of plant species throughout the world, and was probably transmitted to healthy susceptible plants 'through slight abrasive contacts in the presence of viruliferous dusts or of liquid extracts derived from infective materials.' He argued that 'Resistant species or varieties within generally susceptible groups of plants will be most in evidence where a disease has existed longest and thus had an opportunity to discourage highly susceptible types'. He noted (figure 2) that in the genus *Nicotiana*, the species which respond to infection in a hypersensitive manner and would be the poorest reservoirs of infection in nature, are *N. glutinosa*, which is a native of Peru, *N. repanda* of Mexico, *N. rustica* of Ecuador and Peru and *N. langsdorfii of* Brazil, and several species of other genera of the Solanaceae centred in South America, including *Solanum capsicastrum* of Brazil and *S. tuberosum* of Bolivia and Peru. By contrast the *Nicotiana* species that respond to TMV infection with bright chlorosis and mottling, and accumulate the greatest concentrations of virions, are mostly found in North America, southern South America, and Australia. Holmes (1950) concluded that the evidence '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.' He also noted that in this region there are now three species of *Nicotiana* (*N. glauca*, *N. raimondii* and *N. wigandioides*) that tolerate TMV infection, show few or no symptoms and may be the long-term niche of TMV. Incidentally the adaptation of tobamoviruses to contagious transmission by 'abrasive contacts' seems not merely to depend on their ability to produce large numbers of remarkably stable virions. It probably also depends on an unusual host–virion interaction that renders older, bigger and exposed basal leaves more susceptible to infection than the young leaves (see p. 103 of Bawden (1964)); for most

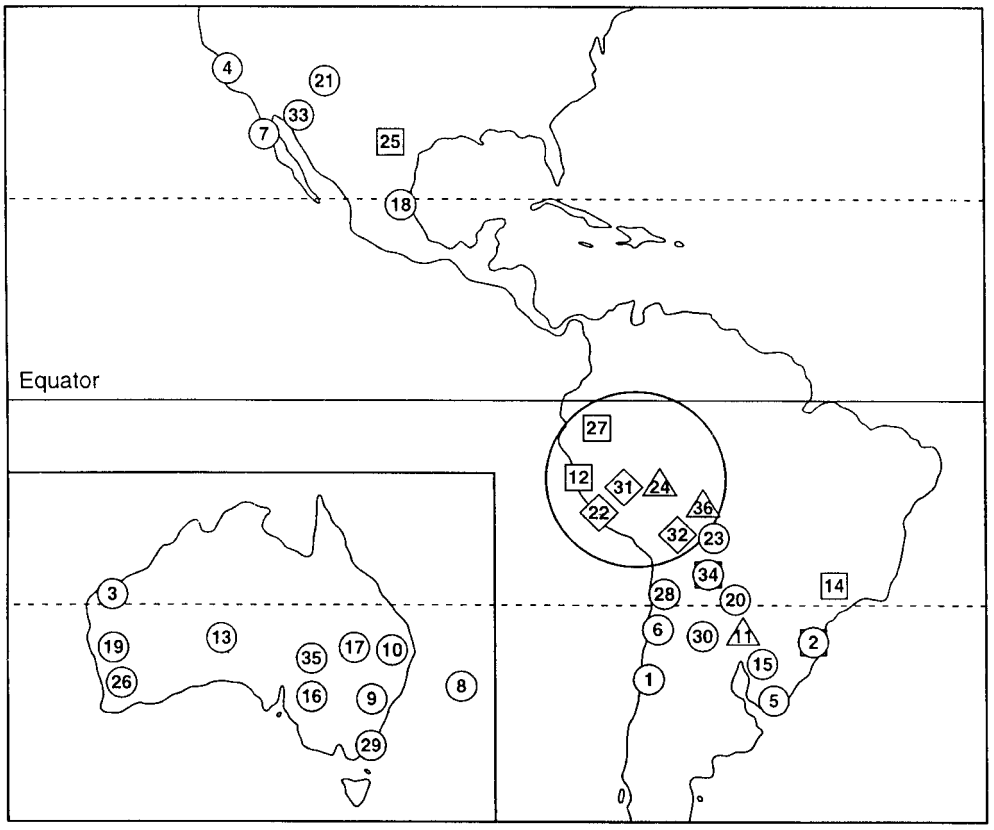

Figure 2. Map of the Americas and Australia (inset) showing, approximately, where *Nicotiana* species, tested for their responses to TMV infection, are endemic (redrawn from Holmes (1950)). Fully susceptible species that show mottling symptoms are circled; hypersensitive species are in squares; species that have some lines hypersensitive and others mottled are in circles within squares; those that are tolerant are in triangles, and species that have restricted infection to small chlorotic areas are in diamonds. Species 1, *N. accuminata*; 2, *N. alata*; 3, *N. benthamiana*; 4, *N. bigelovii*; 5, *N. bonariensis*; 6, *N. caudigera*; 7, *N. clevelandii*; 8, *N. debneyi*; 9, *N. eastii*; 10, *N. exigua*; 11, *N. glauca*; 12, *N. glutinosa*; 13, *N. gossei*; 14, *N. langsdorfii*; 15, *N. longiflora*; 16, *N. maritima*; 17, *N. megalosiphon*; 18, *N. nudicaulis*; 19, *N. occidentalis*; 20, *N. otophora*; 21, *N. palmeri*; 22, *N. paniculata*; 23, *N. plumbaginifolia*; 24, *N. raimondii*; 25, *N. repanda*; 26, *N. rotundifolia*; 27, *N. rustica*; 28, *N. solanifolia*; 29, *N. suaveolens*; 30, *N. sylvestris*; 31, *N. tomentosa*; 32, *N. tomentosiformis*; 33, *N. trigonophylla*; 34, *N. undulata*; 35, N. *velutina*; 36, *N. wigandioides*. Taliansky *et al.* (1994) have shown that TMV induces lesions in *N. megalosiphon* but then spreads systemically.

viruses, and also TMV RNA preparations, the tip leaves are more susceptible than the older ones.

Additional biogeographic and taxonomic clues are too fragmentary at present to add a firm time-scale to Holmes' analysis, although it is likely that millions, not thousands, of years have been involved in the processes that Holmes' analyses imply. The Solanaceae is a mainly tropical family with its earliest fossils known from the Cretaceous, 65 million years ago (D'Arcy 1991). However its distribution shows some Gondwanan features (Symon 1991). Central and South America are its major centre of diversity with minor centres in Eurasia, especially the Himalayan region, and also Australia, and hence it may have originated somewhat earlier; the Indian subcontinent left Gondwana around 80 million years ago (Raven 1983). However sequence analysis of the chloroplast genome of 21 *Nicotiana* species (Olmstead & Palmer 1991), including some studied by Holmes, shows that species hypersensitive to TMV fall into two basal lineages. This implies either that hypersensitivity to TMV is the ancestral condition, or has arisen more than once, or spreads by hybridization between species; an indication, whichever is correct, that the interaction of *Nicotiana* and TMV is ancient. The Australian species are all susceptible to TMV, and gene sequence analysis has shown that they are all closely related to one another and to *N. glauca* and *N. sylvestris*, hence it is likely that they spread to Australia much more recently than the breakup of Gondwana.

Another important finding of the TMGMV studies was that the diversity of that virus population was 'only an order of magnitude greater than that reported for DNA genomes' (Nei 1987, table 4). This causes one to question whether the diversity of all RNA genomes is so qualitatively different from that of DNA genomes as to need a special concept, the 'quasi-species' (Domingo *et al.* 1995), to describe them. The relatively large mutability of RNA genomes does not necessarily produce more diverse populations, it may merely allow them to evolve into a diverse population more quickly (Smith & Inglis 1987).

Much of great theoretical and practical value could be learned from further studies of the coevolution of the tobamovirus lineage that infects solanaceous plants. Holmes' seminal studies could be extended to include other solanaceous tobamoviruses and their hosts, now that so much more is known of the relationships of these viruses and their hosts, and of the molecular biology of hypersensitivity genes (Hammond-Kosack & Jones 1997; Jones & Jones 1997; Parker & Coleman 1997).

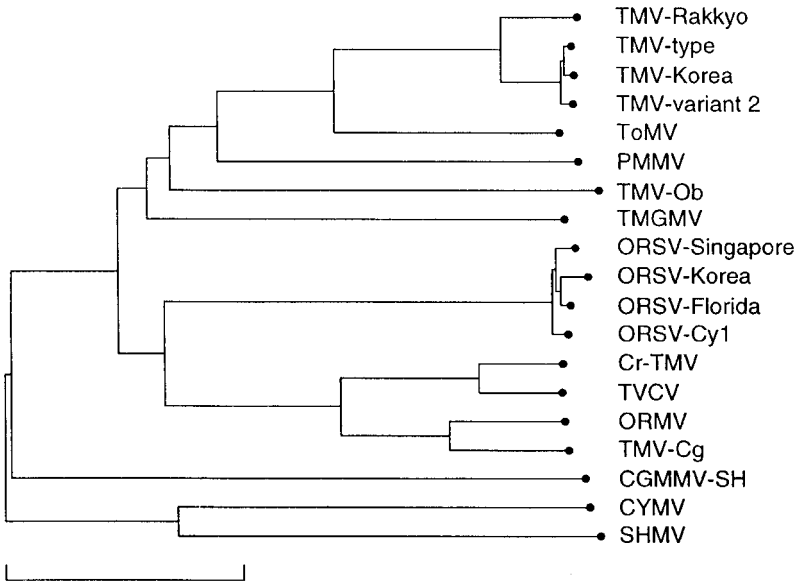

Figure 3. Neighbour-joining tree calculated from the pairwise percentage nucleotide differences (excluding gaps) between the aligned replicase-protein genes of 19 tobamoviruses. Acronyms as in table 1; bar corresponds to a 10% difference (uncorrected for multiple mutations). The sequence of CYMV is, as yet, unpublished (K. Wei, A. M. Mackenzie, and A. J. Gibbs), the others came from the public databases and had the following Accession Codes: CGMMV-SH, D12505; cr-TMV, Z29370; ORMV, U30944; ORSV-Cy1, S83257; ORSV-Florida, U89894; ORSV-Korea, X82130; ORSV-Singapore, U34586; PMMV, M81413; SHMV, J02413; TMGMV, M34077; TMV-Cg, D38444; TMV-Korea, X68110; TMV-Ob, D13438; TMV-Rakkyo, D63809; TMV-Type, J02415; TMV-variant 2, U01409; ToMV, X02144; TVCV, U03387.

Table 1. *The currently recognized tobamovirus species, their acronyms, synonyms and the year in which they were first reported*

1886	tobacco mosaic tobamovirus (TMV): common strain, type strain, U1 strain, *vulgare*;
1909	tomato mosaic tobamovirus (ToMV): *dahlemense*, tomato aucuba mosaic, tomato enation mottle, TMV-strain L
1929	tobacco mild green mosaic tobamovirus (TMGMV): green-tomato atypical mosaic, mild strain of TMV, para-tobacco mosaic, TMV-U2, TMV-U5, tomato atypical mosaic green mottling strain
1935	cucumber green mottle mosaic tobamovirus (CGMMV): TMV-watermelon strain (TMV-W), cucumber virus 3, cucumber virus 4, Indian bottle gourd mosaic, cucumis virus 2
1941	ribgrass mosaic tobamovirus (RMV): TMV-ribgrass strain, Holmes ribgrass, crucifer TMV (cr-TMV), turnip vein clearing (TVCV), TMV-wasabi
1946	sunnhemp mosaic tobamovirus (SHMV): TMV-bean strain, TMV-cowpea strain, cowpea chlorotic spot, cowpea mosaic (now used for a comovirus), cowpea yellow mosaic (now used for a comovirus), *Crotalaria mucronata* mosaic, dolichos enation mosaic, sunnhemp rosette
1951	odontoglossum ringspot tobamovirus (ORSV): TMV-orchid strain
1952	pepper mild mottle tobamovirus (PMMV): capsicum mosaic, paprika mild mottle, pepper mosaic, TMV-Samsun latent strain, also TMV-P8 isolate
1961	Sammons opuntia tobamovirus (SOV)
1967	kyuri green mottle mosaic tobamovirus (KGMMV): Japanese CV3, CGMMV-cucumber strain or strain C
1971	1971 frangipani mosaic tobamovirus (FrMV): temple tree mosaic, champa mosaic
1979	paprika mild mottle tobamovirus (PaMMV)
1982	1982 ullucus mild mottle tobamovirus (UMMV)
1983	1983 tobacco mosaic virus-Obuda (TMV-Ob)
1994	tobacco mosaic virus-Rakkyo (TMV-R);
1996	oilseed rape mosaic tobamovirus (ORMV): Chinese rape mosaic, youcai mosaic, TMV-Cg
1998	clitoria yellow mottle tobamovirus (CYMV)

4. PHYLOGENY OF TOBAMOVIRUSES INFERRED FROM THEIR GENES AND PROTEINS

There were many early reports of different 'strains', 'types' or 'forms' of TMV (Bald 1960; Bald & Goodchild 1960; Broadbent 1962; MacNeill 1962). Some of these isolates had different host preferences, and their relationships were assessed by the serological behaviour (Van Regenmortel 1986), the amino-acid sequence or the composition of their coat proteins (Gibbs 1986). Many have been designated as distinct tobamovirus species, for although they clearly share a common ancestry, they have acquired separate evolutionary histories and occupy different ecological niches.

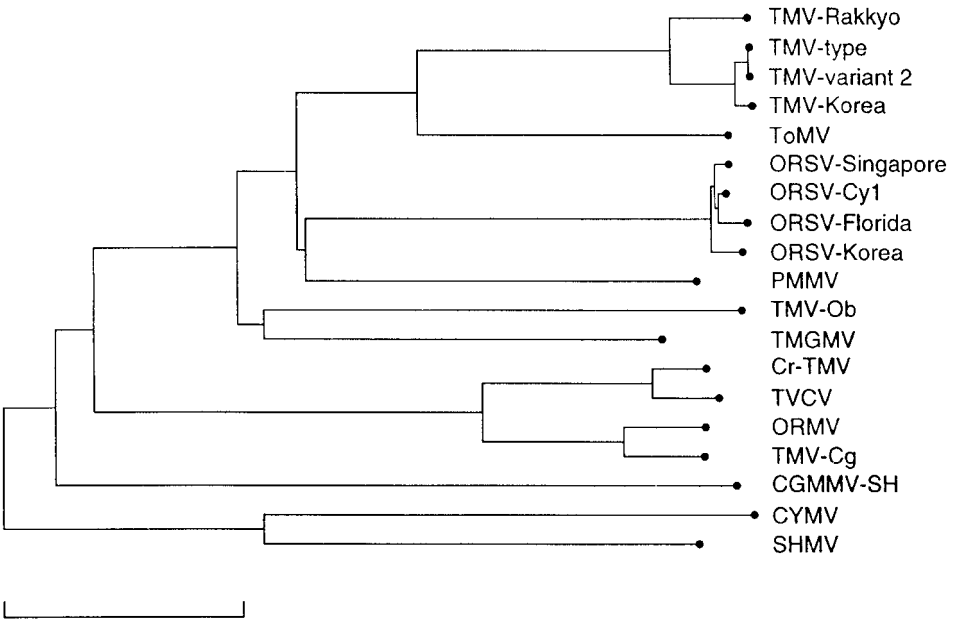

Figure 4. Neighbour-joining tree calculated from the pairwise percentage nucleotide differences (excluding gaps) between the aligned movement-protein genes of 19 tobamoviruses. Acronyms as in table 1; bar corresponds to a 10% difference (uncorrected for multiple mutations). Sequence sources as in figure 3.

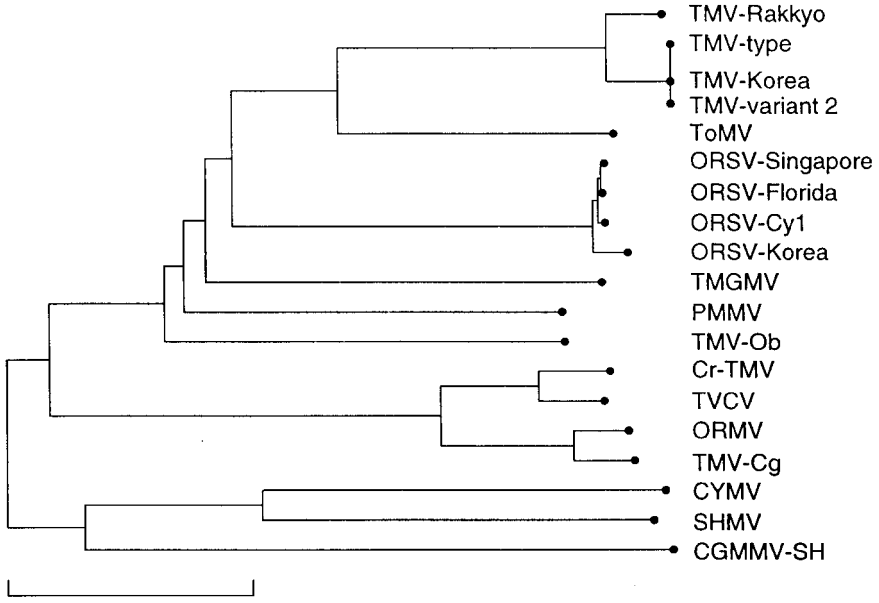

Figure 5. Neighbour-joining tree calculated from the pairwise percentage nucleotide differences (excluding gaps) between the aligned coat protein genes of 19 tobamoviruses. Acronyms as in table 1; bar corresponds to a 10% difference (uncorrected for multiple mutations). Sequence sources as in figure 3.

The nucleotide sequence of the complete genome is known for 19 tobamovirus isolates (mid-1998), together with many more partial sequences, mostly of the coat protein gene. Classifications of these sequences by distance matrix methods (figures 3, 4 and 5) or by parsimony methods (Lartey *et al.* 1996) reveal the same relationships, and largely confirm the earlier studies of the coat proteins mentioned above. The classifications show that the 19 isolates are from perhaps 12 tobamovirus species, which are listed, together with five, other less well-characterized species, in table 1. The relationships between all, except odontoglossum ringspot tobamovirus (ORSV), are closely similar whether the replicase- (figure 3), movement- (figure 4) or coat protein (figure 5) genes or encoded proteins are used for the comparisons; minor differences between these dendrograms are probably insignificant, and could only be resolved by sequencing more genomes. Figure 6*a*,*b* shows a scatter plot, obtained using the DIPLOMO program (Weiller & Gibbs 1995), comparing pairwise nucleotide distances against pairwise amino-acid distances for the aligned sequences of the replicase genes and their encoded proteins. It can be seen that the points closely follow a monotonic, albeit curvilinear, relationship, indicating that this set of aligned sequences is a mostly coherent set of data; only those involving the ORSV genes are atypical (figure 6*b*). The same result was obtained with plots comparing the other genes and their encoded proteins. This confirms the three-way rate tests made by Lartey *et al.* (1996), who showed that all lineages of the tobamoviruses, except ORSV, had remarkably uniform rates of evolution in their replicase and movement-protein genes; there was some evidence of differences among the coat protein genes. The classifications also show that the replicase genes of the ORSVs are closest to those of the brassica tobamoviruses, but their movement- and coat protein genes place them among those isolated from Solanaceae, thus the ancestral

ORSV was, most probably, a recombinant (Lartey *et al*. 1996; Gibbs *et al*. 1997).

The major groupings of tobamoviruses calculated from their genomic sequences (figures 3–5) correlate well with groupings based on other criteria. Fukuda *et al*. (1980) noted that there are two major groups of tobamoviruses. Those infecting cucurbits and legumes (CGMMV, SHMV and CYMV) form one group, characterized by having their origin of virion assembly region within the coat protein gene, and hence also within the mRNA for the coat protein. Whereas the virion assembly region of all others is within the movement-protein gene. These lineages coincide with two major lineages of the nucleotide sequence classification. This grouping is further reinforced by the conclusion of Lartey *et al*. (1996) that the brassica infecting tobamoviruses form a distinct sublineage of the tobamoviruses characterized by having replaced a small deleted 3′-terminal portion of their movement-protein gene by 'overprinting' the 5′-terminal part of their coat protein gene. Lartey *et al*. (1996) deduced from the overprinting that the brassica-infecting tobamoviruses were the 'derived', rather than the 'ancestral', group. Finally, the structure of the coat proteins of five tobamoviruses (Wang & Stubbs 1994; Wang *et al*. 1997, 1998) is known and clearly confirms the pattern of genomic sequence relationships; the TMV, TMGMV and ORSV proteins are most similar, but distinct from that of the RMV protein, and the most distinct of all is the coat protein of CGMMV.

It would, of course, be of great interest to estimate the age of the tobamoviruses; namely to place a time-scale on the ordinates of figures 3–5. Nearly 20 years ago I attempted to answer the question, 'How ancient are the tobamoviruses?' (Gibbs 1980), by extrapolation. The coat protein sequences of seven tobamoviruses known at that time, were maximally about 120% different in sequence, after correction for multiple mutations. If they were evolving at a rate of 1% every 2–10 million years, like many other proteins (Wilson *et al*. 1977), then the proto-tobamovirus probably arose about 120–600 million years ago. However a safer way to infer the age of a fossil-less group of organisms, like the tobamoviruses, is to link their phylogeny with that of a group whose fossil history is known, in this instance their hosts. Thus, it is significant that the 19 genome sequences fall into lineages that mostly correlate with the family of the host from which they were isolated; a relationship not found with most other virus genera. One large cluster of tobamoviruses includes all those isolated from solanaceous plants, other pairs (on long branches) infect brassicas and legumes, and similarly distinct are those isolated from cucurbits and orchids. No gene sequence data are available for kyuri green mild mosaic virus, which is the other distinct tobamovirus isolated from cucurbits (Francki & Palukaitis 1986), but peptide mapping of its replicase protein (Fraile & Garcia-Arenal 1990) has shown that its relationship to CGMMV is similar to that between SHMV and CYMV. A major division of the eudicotyledonous angiosperms, first described by Young & Watson (1970) and broadly confirmed by RubisCo sequence comparisons (Donoghue 1998), is that which separates the more ancestral crassinucellate plants (Rosidae, Caryophyllidae and others) from the clade of tenuinucellate plants (Asteridae). The cucurbits and legumes are crassinucellate, and the Solanaceae are tenuinucellate. Thus, the higher-level relationships of the hosts correlate well with the pattern of relationships of the major tobamovirus groups.

The simplest explanation for this correlation between host and virus grouping is that most of the tobamoviruses have coevolved with their hosts (Gibbs 1986; Lartey *et al*. 1996); other scenarios, involving movement between host lineages and selection that would make them appear similar, provide less parsimonious explanations. Furthermore, the fact that tobamovirus phylogenetic trees calculated for synonymous nucleotide differences are the same as those given by all differences indicates that selection of the encoded proteins is not providing the phylogenetic signals. Thus, it seems likely that the long-term survival of most tobamovirus species in nature depends on their adaptation to a particular family of plants. Tobamoviruses are often isolated from unexpected sources, and although some of these associations persist, like that of ORSV and orchids, and of TMV-Rakkyo and *Allium chinense*, a chives-like herb used in China and Japan (Chen *et al*. 1996), most do not.

The likelihood that the tobamoviruses have codiverged with their hosts indicates the time-scale of their relationships. The angiosperms arose 120–140 million years ago and most modern families had evolved before 60–80 million years ago (Raven 1983). Thus, the root of the tobamovirus phylogeny, which is probably between or within the CGMMV and SHMV lineages, may correspond to proto-tobamoviruses infecting the earliest angiosperms. The major tobamovirus radiations resulting in the clusters now found in the Solanaceae, the legumes and cucurbits may have occurred when these modern families radiated 60–80 million years ago. This period included both the final stages of dismemberment of Gondwana, and also the Cretaceous–Tertiary extinction boundary associated with the Chicxulub impact (Schultz & Dhondt 1996), events that may also account for the deep branches within some of these tobamovirus groups. The legume-infecting tobamovirus, SHMV, seems not to occur naturally in Australia, whereas CYMV has only been found in Australia, and similarly CGMMV is found worldwide, whereas KGMMV has only been found in Japan. Perhaps these pairs have been separated by continental drift since the end of the Cretaceous Period.

At variance with the idea that the tobamoviruses and their hosts have coevolved is the fact that those isolated from brassicas, which are crassinucellate, are most closely related to, and derived from, the tenuinucellate tobamoviruses. However, at least one of the brassica-infecting tobamoviruses is best known as ribgrass mosaic virus (Holmes 1941) as it is common throughout the world in species of *Plantago*, a tenuinucellate genus, and hence the long-term hosts of the brassica-infecting tobamoviruses may be plantains. Thus, the ability to prosper in both brassicas and plantains, like the gene sequences encoding their movement- and/or coat protein junction, may be a 'derived' rather than an 'ancestral' state. ORSV also has an anomalous host range, but is a recombinant. Lartey *et al*. (1996) noted that ORSV has a significantly faster rate of evolution, and suggested that this might be

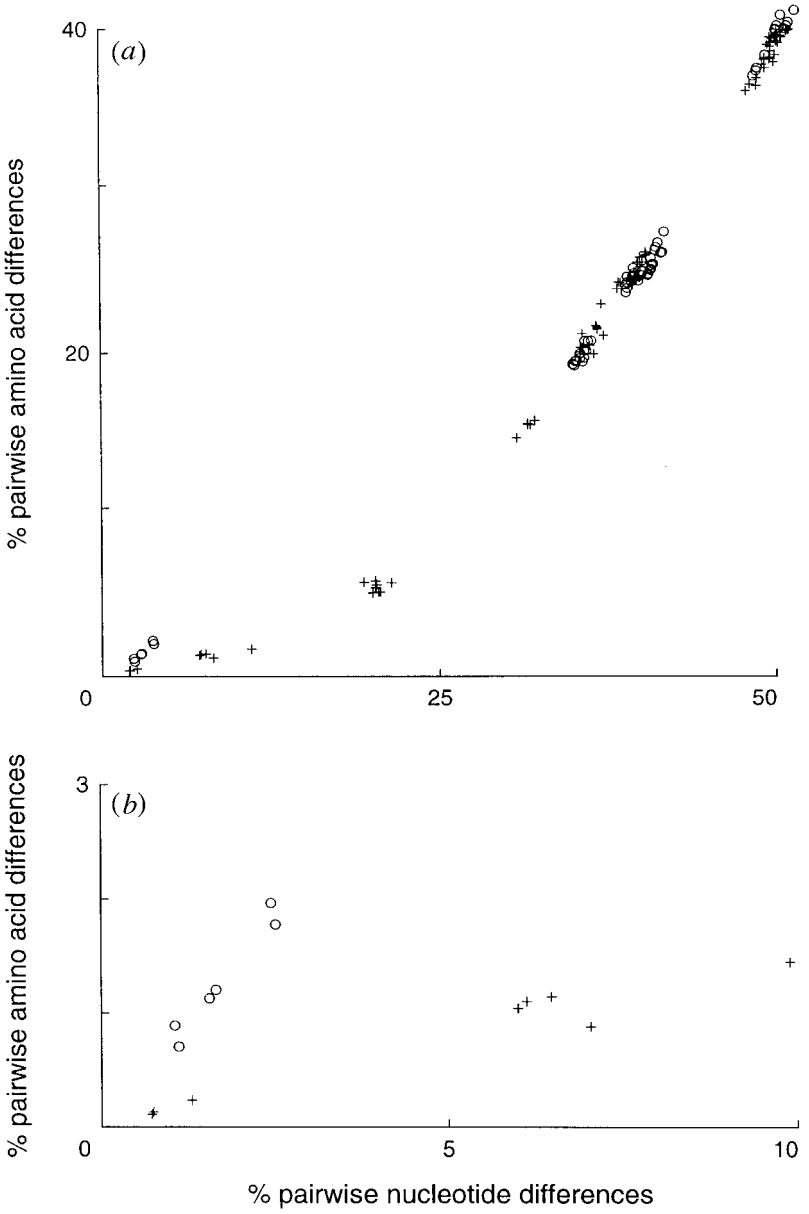

Figure 6. Scatter plot of the percentage nucleotide difference (excluding gaps) on the *x*-axis against the corresponding percentage amino-acid difference on the *y*-axis, for each pairwise comparison of 19 tobamoviruses, obtained using the DIPLOMO program. Comparisons involving ORSV as one or both of the pair are shown as circles, others as '+'. (*a*) The full graph; (*b*) the lower left corner of (*a*) enlarged.

a concomitant of recombination. However the long branches linking the different genes of ORSV either to the brassica-infecting or to the solanaceous tobamoviruses in figures 3–5 suggest that the recombinational event which produced ORSV occurred long ago, whereas its significantly larger amino acid:nucleotide ratio (figure 6*a*,*b*) is probably more recent. Thus, its present monocotyledonous hosts may have been acquired recently and have positively selected amino-acid changes, especially in its replicase, which has a greater shift in amino acid:nucleotide ratio than the other genes.

ORSV is not the only tobamovirus to provide evidence of past genetic recombination, which has clearly been as important a feature of the evolution of some tobamoviruses as it is of viruses in other groups. Fraile *et al.* (1997) found that a tobamovirus isolate obtained from a herbarium specimen of *N. glauca* had the nucleotide sequence of a TMV strain between nucleotides 912 and 1250, and of TMGMV between nucleotides 3461 and 3769. Furthermore Lartey *et al.* (1996) found possible evidence of recombination in the middle of the TMGMV and TMV-Ob genomes. However the conclusion by Meshi *et al.* (1981) that the tRNA-like sequence at the 3′-terminus of the SHMV genome might have been acquired by recombination from a tymovirus is probably incorrect. The SHMV tRNA-like sequence accepts valine, rather than histidine like most tobamoviruses (Hall 1979), however its tRNA-like sequence, together with that of CYMV, is no closer to those of tymoviruses and other tobamoviruses than those of some furoviruses (Goodwin & Dreher 1998).

5. ORIGINS OF TOBAMOVIRUS GENES

Before methods to determine gene sequences were invented in the 1970s, there was little but speculation to promote discussion of the origins and modes of evolution of viruses. Virus groups had been defined using characters such as host type, virion morphology and serological specificity, but the relationships, if any, between such groups were unknown.

The discovery of the GDD-sequence motif in a wide range of viral polymerases (Kamer & Argos 1984; Argos 1988) showed that viruses, previously thought be to unrelated, had genes that appeared to be related. Haseloff *et al*. (1984) first showed that four different viruses, one of them TMV, not known to be related in any way other than that their genomes were single-stranded RNA, had clearly related replication proteins. Most surprising was the fact that three of the viruses only replicated in plants whereas the fourth, Sindbis alphavirus, replicated only in vertebrates and invertebrates! Haseloff and his colleagues speculated that 'Reassortment of functional modules of coding and regulatory sequence from preexisting viral or cellular sources, perhaps via RNA recombination, may be an important mechanism in RNA virus evolution.' Thus, the process of 'modular evolution', first detected in bacteriophages (Botstein 1980), emerged as a major feature of the evolution of all viruses.

Subsequently, there has been a flood of reports (Bruenn 1991; Gorbalenya *et al*. 1990; Habili & Symons 1989; Koonin & Dolja 1993; Morozov *et al*. 1989; Mushegian & Koonin 1993) of sequence motifs in viral proteins unexpectedly suggesting linkages between different higher taxa (Zaccomer *et al*. 1995). However, use of simple robust Monte-Carlo sequence shuffling tests (Melcher 1990; Zanotto *et al*. 1996) has cast doubt on the reality of most of the more distant groupings claimed. For example Koonin & Dolja (1993) used motifs to infer 'that all positive-strand RNA viruses and some related double-stranded RNA viruses could have evolved from a common ancestor virus that contained genes for RNA-dependent RNA polymerase, a chymotrypsin-related protease that also functioned as the capsid protein, and possibly an RNA helicase!' By contrast, Zanotto *et al*. (1996) using Monte-Carlo analyses placed the same polymerases in at least a dozen separate groups and found no statistical evidence of relatedness between those groups.

Nonetheless, both approaches (Koonin & Dolja 1993; Zanotto *et al*. 1996) agree that the tobamoviruses share their RNA polymerase genes with other species of a large group of viruses called the 'alpha-like' virus group (ALVG) by Goldbach & de Haan (1994). This group has several lineages. In the 'tobamo sub-group', the tobamoviruses are closest to hordeiviruses and tobraviruses and also soilborne wheat mosaic virus, and this cluster forms a sister group to other plant-infecting genera that include alfamoviruses, cucumoviruses, bromoviruses, idaeoviruses and ilarviruses. The 'tobamo sub-group' is a sister group to the 'alpha sub-group' which includes the animal-infecting alphaviruses, rubiviruses, hepatitis E virus and beet necrotic yellow vein virus (Koonin & Dolja 1993). The 5′-terminal methyl transferases and helicases of the tobamoviruses also show the same relationships (Koonin & Dolja 1993), indicating that these polymerases, transferases and helicases have formed a module with a long phylogenetic history.

By contrast, the ancestry of the tobamovirus 30 K movement protein is uncertain. Those of different tobamoviruses are clearly related to one another (Melcher 1990), but the suggestion that the movement proteins of different genera form larger groupings (Mushegian & Koonin 1993) is not convincing. The movement proteins of plant viruses are probably polyphyletic in origin. One possible source of these genes that has not been fully explored is that they are overprinted genes (Gibbs & Keese 1995), that have arisen *de novo* during the establishment of each genus; the slight similarity of motifs found in some of them may reflect selection for similar function or composition rather than shared ancestry.

The tobamovirus coat proteins are also clearly related in sequence and structure (Dolja *et al*. 1991) to those of other viruses with rod-shaped and filamentous virions, especially the tobraviruses (Goulden *et al*. 1992) and also the hordeiviruses and furoviruses. The core of these proteins consists of a bundle of four alpha helices, which is a structural fold found in many other proteins, and which shows evidence of having arisen by duplication of a two-helix protein (McLachlan *et al*. 1980).

Thus, all genomes of the ALVG share a gene module encoding a methyl transferase, helicase and polymerase, and some of them, notably the tobamoviruses and related viruses with rod-shaped virions also share the coat protein gene. In the hordeiviruses, idaeoviruses, tobraviruses and tricornaviruses (alfamoviruses, bromoviruses, cucumoviruses and ilarviruses) the enzyme module has become divided and the corresponding genes now occur on separate genome segments. Also, in the 'alpha sub-group', a papain-like serine protease gene has been inserted between the methyl transferase and helicase genes.

If it is correct that the tobamoviruses coevolved with the angiosperms, then the links with other viruses of the ALVG occurred before 120–140 million years ago. More evidence of the pre-tobamovirus lineages is coming from a tobamo-like virus isolated from the alga *Chara australis* (Gibbs *et al*. 1975; Skotnicki *et al*. 1976). The virions of *C. australis* virus (CAV) closely resemble those of TMV, but are 530 nm long, and react distantly with high-titred antisera to some tobamoviruses, as too do those of soil-borne wheat mosaic and potato mop top viruses (Kassanis *et al*. 1972; Powell 1976). At present (July 1998) over 9 kb of the genome of this virus has been sequenced (P. Keese, A. M. Mackenzie, M. Torronen and A. J. Gibbs, unpublished data). Database searches show that the coat protein of CAV is most closely related to that of CGMMV, although more distantly than the coat proteins of tobamoviruses are related to one another. The remainder of the genome is a mosaic of segments that best match, although distantly, either parts of the genome of beet necrotic yellow vein virus or of hepatitis E virus, or that match nothing in current databases.

Beet necrotic yellow vein and hepatitis E viruses are from the 'alpha sub-group' of the ALVG. Thus, CAV is perhaps the first described species of a third lineage of this group, and has genetic links with both of its major lineages. It is possible that the ALVG and charophycean algae have had a very long-term association as these algae are the probable ancestors of land plants (Kranz *et al*. 1995); the earliest charalean fossils are from late Silurian–Devonian period 420 million years ago (Tappan 1980; Grant 1989). Interestingly, although the mode of transmission of CAV is not known, plasmodiophoromycete fungi are known to parasitize charalean algae (Karling 1968), and other species transmit the furoviruses of the ALVG, including soil-borne wheat mosaic, peanut clump and potato mop top viruses. It would be of great interest to find and study more ALVG viruses

of charophycean alga, especially dioecious charalean algae, as it seems that their biogeography reflects plate tectonics (Proctor 1980).

I am indebted to Fernando Garcia-Arenal and Mark Gibbs for very helpful criticisms of a draft of this paper.

REFERENCES

Aldaoud, R., Dawson, W. O. & Jones, G. E. 1989 Rapid, random evolution of the genetic structure of replicating tobacco mosaic virus populations. *Intervirology* **30**, 227–233.

Argos, P. 1988 A sequence motif in many polymerases. *Nucl. Acids Res.* **16**, 9909–9916.

Bald, J. G. 1960 Forms of tobacco mosaic virus. *Nature* **188**, 645–647.

Bald, J. G. & Goodchild, D. J. 1960 Tobacco mosaic virus in *Nicotiana glauca*. *Phytopathology* **50**, 497–499.

Bawden, F. C. 1964 *Plant viruses and virus diseases*, 4th edn. New York: Ronald Press.

Botstein, D. 1980 A theory of modular evolution for bacteriophages. *Ann. NY Acad. Sci.* **354**, 484–491.

Broadbent, L. 1962 The epidemiology of tomato mosaic. II. Smoking tobacco as a source of virus. *Ann. Appl. Biol.* **50**, 461–466.

Bruenn, J. A. 1991 Relationships among the positive strand and double-strand RNA viruses as viewed through their RNA-dependent RNA polymerases. *Nucl. Acids Res.* **19**, 217–226.

Chen, J., Watanabe, Y., Sako, N., Oshima, K. & Okada, Y. 1996 Mapping of host range restriction of the rakkyo strain of tobacco mosaic virus in *Nicotiana tabacum* cv. bright yellow. *Virology* **226**, 198–204.

D'Arcy, W. G. 1991 The Solanaceae since 1976, with a review of its biogeography. In *Solanaceae. III. Taxonomy, chemistry, evolution* (ed. J. G. Hawkes, R. N. Lester, M. Nee & N. Estrada), pp. 75–137. London: Royal Botanic Gardens and Linnaean Society of London.

Dawson, W. O., Beck, D. L. & Knorr, D. A. 1986 cDNA cloning of the complete genome of tobacco mosaic virus and production of infectious transcripts. *Proc. Natl Acad. Sci. USA* **83**, 1832–1836.

Dolja, V. V., Boyko, V. P., Agranovsky, A. A. & Koonin, E. V. 1991 Phylogeny of capsid proteins of rod-shaped and filamentous RNA plant viruses: two families with distinct patterns of sequence and probably structure conservation. *Virology* **184**, 79–86.

Domingo, E., Holland, J., Biebricher, C. & Eigen, M. 1995 Quasi-species: the concept and the word. In *Molecular basis of virus evolution* (ed. A. J. Gibbs, C. H. Calisher & F. Garcia-Arenal) pp. 181–191. Cambridge University Press.

Donoghue, M. J. 1998 Angiosperms. At: http://phylogeny.arizona.edu/tree/eukaryotes/green.plants/embryophytes/angiosperms/angiosperms.html.

Fraile, A. & Garcia-Arenal, F. 1990 A classification of the tobamoviruses based on comparisons among their 126 K proteins. *J. Gen. Virol.* **71**, 2223–2228.

Fraile, A., Malpica, J. M., Aranda, M. A., Rodriguez-Cerezo, E. & Garcia-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. & Garcia-Arenal, F. 1997 A century of tobamovirus evolution in an Australian population of *Nicotiana glauca*. *J. Virol.* **71**, 8316–8320.

Francki, R. I. B. & Palukaitis, P. 1986 Taxonomy of cucurbit-infecting tobamoviruses as determined by serological and molecular hybridization analyses. *Intervirology* **26**, 156–163.

Fukuda, M., Okada, Y., Otsuki, Y. & Takebe, I. 1980 The site of initiation of rod assembly on the RNA of a tomato and a cowpea strain of tobacco mosaic virus. *Virology* **101**, 493–502.

Gibbs, A. 1980 How ancient are the tobamoviruses? *Intervirology* **14**, 101–108.

Gibbs, A. 1986 Tobamovirus classification. In *The plant viruses. 2. The rod-shaped plant viruses* (ed. M. H. V. van Regenmortel & H. Fraenkel-Conrat), pp. 167–180. New York: Plenum Press.

Gibbs, A. & Keese, P. 1995 In search of the origins of viral genes. In *Molecular basis of virus evolution* (ed. A. J. Gibbs, C. H. Calisher & F. Garcia-Arenal), pp. 76–90. Cambridge University Press.

Gibbs, A., Skotnicki, A. H., Gardiner, J. E. & Walker, E. S. 1975 A tobamovirus of a green alga. *Virology* **64**, 571–574.

Gibbs, M. J., Armstrong, J., Weiller, G. F. & Gibbs, A. J. 1997 Virus evolution; the past a window on the future? In *Virus-resistant plants: potential ecological impact* (ed. M. Tepfer & E. Balazs), pp. 1–19. Berlin: Springer.

Goldbach, R. & de Haan, P. 1994 RNA viral supergroups and the evolution of RNA viruses. In *The evolutionary biology of viruses* (ed. S. Morse) pp. 161–184. New York: Raven Press.

Goodwin, J. B. & Dreher, T. W. 1998 Transfer RNA mimicry in a new group of positive-strand RNA plant viruses, the furoviruses: differential aminoacylation between the RNA components of one genome. *Virology* **246**, 170–178.

Gorbalenya, A. E., Koonin, E. V. & Wolf, Y. I. 1990 A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. *FEBS Lett.* **262**, 145–148.

Goulden, M. G., Davies, J. W., Wood, K. R. & Lomonossoff, G. P. 1992 Structure of tobraviral particles: a model suggested from sequence conservation in tobraviral and tobamoviral coat proteins. *J. Mol. Biol.* **227**, 1–8.

Grant, M. C. 1989 Phylum Chlorophyta; Class Charophyceae; Order Charales. In *Handbook of Protoctista* (ed. L. Margulis, J. O. Corliss, M. Melkonian & D. J. Chapman), pp. 641–648. Boston, MA: Jones & Bartlett.

Habili, N. & Symons, R. H. 1989 Evolutionary relationship between luteoviruses and other RNA plant viruses based on sequence motifs in their putative RNA polymerases and nucleic acid helicases. *Nucl. Acids Res.* **17**, 9543–9555.

Hahn, B. H., Shaw, G. M., Taylor, M. E., Redfield, R. R., Markham, P. D., Salahuddin, S. Z., Wong-Staal, F., Gallo, R. C., Parks, E. S. & Parks, W. P. 1986 Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk from AIDS. *Science* **232**, 1548–1553.

Hall, T. C. 1979 Transfer RNA-like structures in viral genomes. *Int. Rev. Cytol.* **60**, 1–26.

Hammond-Kosack, K. E. & Jones, J. D. G. 1997 Plant disease resistance genes. *A. Rev. Plant Physiol.* **48**, 575–607.

Haseloff, J., Goelet, P., Zimmern, D., Ahlquist, P., Dasgupta, R. & Kaesberg, P. 1984 Striking similarities in amino acid sequence among nonstructural proteins encoded by RNA viruses that have dissimilar genomic organization. *Proc. Natl Acad. Sci. USA* **81**, 4358–4362.

Holmes, F. O. 1934 A masked strain of tobacco-mosaic virus. *Phytopathology* **24**, 845–873.

Holmes, F. O. 1941 A distinctive strain of tobacco-mosaic virus from *Plantago*. *Phytopathology* **31**, 1089–1098.

Holmes, F. O. 1950 Indications of a New-World origin of tobacco-mosaic virus. *Phytopathology* **41**, 341–349.

Johnson, J. 1947 Virus attenuation and the separation of strains by specific hosts. *Phytopathology* **37**, 822–837.

Jones, D. A. & Jones, J. D. G. 1997 The role of leucine-rich repeat proteins in plant defences. *Adv. Bot. Res.* **24**, 89–167.

Jones, G. E. & Dawson, W. O. 1978 Stability of mutations conferring temperature sensitivity on tobacco mosaic virus. *Intervirology* **9**, 149–155.

Kamer, G. & Argos, P. 1984 Primary structural comparisons of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucl. Acids Res.* **12**, 7269–7282.

Karling, J. S. 1968 *The Plasmodiophorales*. New York: Hafner.

Kassanis, B. 1957 Some effects of varying temperature on the quality and quantity of tobacco mosaic virus in infected plants. *Virology* **4**, 187–199.

Kassanis, B., Woods, R. D. & White, R. F. 1972 Some properties of potato mop-top virus and its serological relationships to tobacco mosaic virus. *J. Gen. Virol.* **14**, 123–132.

Koonin, E. V. & Dolja, V. V. 1993 Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit. Rev. Biochem. Mol. Biol.* **28**, 375–430.

Kranz, H. D., Miks, D., Siegler, M. L., Capesius, I., Sensen, C. W. & Huss, V. A. 1995 The origin of land plants: phylogenetic relationships among charophytes, bryophytes, and vascular plants inferred from complete small-subunit ribosomal RNA gene sequences. *J. Mol. Evol.* **41**, 74–84.

Lartey, R. T., Voss, T. C. & Melcher, U. 1996 Tobamovirus evolution: gene overlaps, recombination, and taxonomic implications. *Mol. Biol. Evol.* **13**, 1327–1338.

Li, W.-H. 1993 Unbiased estimates of the rates of synonymous and nonsynonymous substitutions. *J. Mol. Biol.* **36**, 96–99.

Lynch, M. R., Burger, R., Butcher, D. & Gabriel, W. 1993 The mutational meltdown in asexual populations. *J. Hered.* **84**, 339–344.

McKinney, H. H. 1935 Evidence of virus mutation in the common mosaic of tobacco. *J. Agric. Res.* **51**, 951–981.

McLachlan, A. D., Bloomer, A. C. & Butler P. J. G. 1980 Structural repeats and evolution of tobacco mosaic virus coat protein and RNA. *J. Mol. Biol.* **136**, 203–224.

MacNeill, B. H. 1962 A specialized tomato form of tobacco mosaic virus in Canada. *Can. J. Bot.* **40**, 49–51.

Melcher, U. 1990 Similarities between putative transport proteins of plant viruses. *J. Gen. Virol.* **71**, 1009–1018.

Meshi, T., Ohno, T., Iba, H. & Okada, Y. 1981 Nucleotide sequence of a cloned cDNA copy of TMV (cowpea strain) RNA including the assembly origin, the coat protein cistron, and the 3′ non-coding region. *Mol. Gen. Genet.* **184**, 20–25.

Morozov, S. Y., Dolja, V. V. & Atabekov, J. G. 1989 Probable reassortment of genomic elements among elongated RNA-containing plant viruses. *J. Mol. Evol.* **29**, 52–62.

Moya, A., Rodriguez-Cerezo, E. & Garcia-Arenal, F. 1993 Genetic structure of natural populations of the plant RNA virus tobacco mild green mosaic virus. *Mol. Biol. Evol.* **10**, 449–456.

Mushegian, A. R. & Koonin, E. V. 1993 Cell-to-cell movement of plant viruses. *Arch. Virol.* **133**, 239–257.

Nei, M. 1987 *Molecular evolutionary genetics*. New York: Columbia University Press.

Olmstead, R. & Palmer, J. D. 1991 Chloroplast DNA and systematics of the Solanaceae. In *Solanaceae. III. Taxonomy, chemistry, evolution* (ed. J. G. Hawkes, R. N. Lester, M. Nee & N. Estrada), pp. 161–168. London: Royal Botanic Gardens and Linnaean Society of London.

Parker, J. E. & Coleman, M. J. 1997 Molecular intimacy between proteins specifying plant–pathogen recognition. *Trends Biochem. Sci.* **22**, 291–296.

Powell, C. A. 1976 The relationship between soil-borne wheat mosaic virus and tobacco mosaic virus. *Virology* **71**, 453–462.

Proctor, V. W. 1980 Historical biogeography of *Chara* (Charophyta): an appraisal of the Braun-Wood classification plus a falsifiable alternative for future consideration. *J. Phycol.* **16**, 218–233.

Raven, P. H. 1983 The migration and evolution of floras in the southern hemisphere. *Bothalia* **14**, 325–328.

Rodriguez-Cerezo, E., Moya, A. & Garcia-Arenal, F. 1989 Variability and evolution of the plant RNA virus pepper mild mottle virus. *J. Virol.* **63**, 2198–2203.

Rodriguez-Cerezo, E., Elena, S. F., Moya, A. & Garcia-Arenal, F. 1991 High genetic stability in natural populations of the plant RNA virus tobacco mild green mosaic virus. *J. Mol. Evol.* **32**, 328–332.

Schultz, P. H. & Dhondt, S. 1996 Cretaceous–Tertiary (Chicxulub) impact angle and its consequences. *Geology* **24**, 963–967.

Skotnicki, A., Gibbs, A. & Wrigley, N. G. 1976 Further studies on *Chara corallina* virus. *Virology* **75**, 457–468.

Smith, D. B. & Inglis, S. C. 1987 The mutation rate and variability of eukaryotic viruses: an analytical review. *J. Gen. Virol.* **68**, 2729–2740.

Symon, D. E. 1991 Gondwanan elements of the Solanaceae. In *Solanaceae. III. Taxonomy, chemistry, evolution* (ed. J. G. Hawkes, R. N. Lester, M. Nee & N. Estrada), pp. 139–150. London: Royal Botanic Gardens and Linnaean Society of London.

Taliansky, M., Aranda, M. A. & Garcia-Arenal, F. 1994 Differential invasion by tobamoviruses of *Nicotiana megalosiphon* following the hypersensitive response. *Phytopathology* **84**, 812–815.

Tappan, H. 1980 *The paleobiology of plant protists*. San Fransisco: W. H. Freeman.

Van Regenmortel, M. H. V. 1986 Tobacco mosaic virus: antigenic structure. In *The plant viruses. 2. The rod-shaped plant viruses* (ed. M. H. V. van Regenmortel & H. Fraenkel-Conrat), pp. 79–104. New York: Plenum Press.

Wang, H. & Stubbs, G. 1994 Structure determination of cucumber green mottle mosaic virus by X-ray fiber diffraction. *J. Mol. Biol.* **239**, 371–384.

Wang, H. Culver, J. N. & Stubbs, G. 1997 Structure of ribgrass mosaic virus at 2.9 Å resolution: evolution and taxonomy of tobamoviruses. *J. Mol. Biol.* **269**, 769–779.

Wang, H., Planchart, A. & Stubbs, G. 1998 Caspar carboxylates: the structural basis of tobamovirus disassembly. *Biophys. J.* **74**, 633–638.

Weiller, G. F. & Gibbs A. 1995 DIPLOMO: the tool for a new type of evolutionary analysis. *CABIOS* **11**, 535–540.

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

Yarwood, C. E. 1979 Host passage effects with plant viruses. *Adv. Virus Res.* **25**, 169–190.

Young, D. J. & Watson, L. 1970 The classification of dicotyledons: a study of the upper levels of the hierarchy. *Austr. J. Bot.* **18**, 387–433.

Zaccomer, B., Haenni, A.-L. & Macaya, G. 1995 The remarkable variety of plant RNA virus genomes. *J. Gen. Virol.* **76**, 231–247.

Zanotto, P. M. de A., Gibbs, M. J., Gould, E. A. & Holmes, E. C. 1996 A reevaluation of the higher taxonomy of viruses based on RNA polymerases. *J. Virol.* **70**, 6083–6096.