Viruses – evolving organisms?

D.J. Robinson

Ever since viruses were first discovered, there has been argument as to whether or not they are organisms. Beijerinck is often considered to be the "father" of virology because in 1898 he correctly interpreted the observation that the agent of tobacco mosaic disease could pass through a sterilizing filter as showing that the agent was fundamentally different from previously known disease-causing micro-organisms. In contrast, Ivanovskii, who had carried out the same experiment as early as 1892, was still arguing in 1903 that the agent was simply a very small bacterium. As knowledge of the properties of viruses grew, it became evident that Beijerinck's view was the correct one. Viruses lack many of the properties that are regarded as characteristic of organisms: viruses are not composed of cells, they have no metabolism, they do not grow (*i.e.* they do not get larger as they get older), and indeed on their own they are lifeless. By 1928, chapters on viruses were appearing in textbooks on colloid chemistry, and the crystallization of tobacco mosaic virus by Stanley in 1935 seemed to many finally to demonstrate that viruses are "merely" chemicals, albeit very complex ones. However, it was also evident that in a suitable host cell, viruses alter the metabolism of that cell, replicate and multiply, and that they have genetic systems that are comparable to those of organisms. Thus by 1937, Delbrück and others had realized that viruses were good model systems in which to study replication and mutation. For pathologists, an important consequence of this organism-like behaviour is that viruses evolve in much the same ways as do organisms.

In this article, I will describe some of the characteristics of virus evolution, using examples drawn from work done at SCRI.

New forms appear

The essential feature of evolution is that new forms appear that have not existed previously. For higher organisms, this occurs over long periods of time, but because viruses replicate very rapidly, recognizably novel forms appear on an observable time scale. Some of the best examples of the appearance of new forms of virus are among the begomoviruses. Viruses in the genus *Begomovirus* infect dicotyledonous plants and are transmitted by the whitefly, *Bemisia tabaci*. Their genomes are circular single-stranded DNA, and most have two genome parts, DNA-A and DNA-B, each of about 2.8 kb. All functions required for replication, control of gene expression and encapsidation are encoded on DNA-A, and genes involved in intra- and intercellular movement are on DNA-B (Fig. 1). Begomoviruses occur worldwide in tropical and warm temperate regions, and cause many diseases of crops and wild plants.

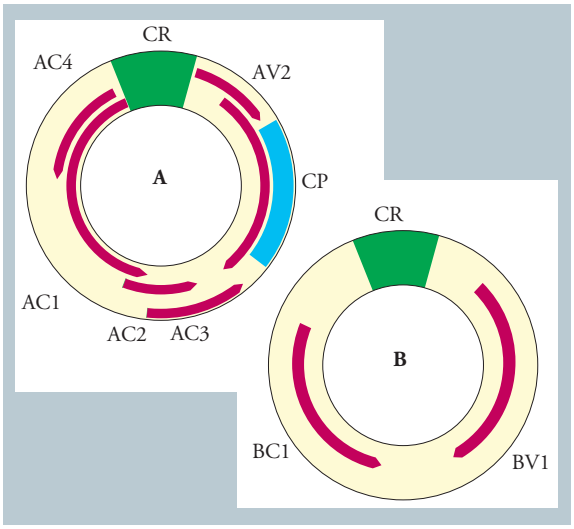

Figure 1 Diagram of the genome of a typical begomovirus. Arrows indicate the positions of genes: CP is the coat protein gene. The region marked in blue is the region derived from ACMV in the recombinant Uganda variant.

Cassava mosaic disease in Uganda. Mosaic diseases of cassava are an important constraint on the production of this staple crop throughout Africa and the Indian sub-continent. We and others have identified five distinct begomovirus species that cause essentially similar diseases: African cassava mosaic virus (ACMV), East African cassava mosaic virus (EACMV), South African cassava mosaic virus, Indian cassava mosaic virus and Sri Lankan cassava mosaic virus. Although earlier work suggested that these viruses each occurred in a different geographical region (see *Annual Report for 1990*, pp.88-90), it is now clear that there is some overlap in their distributions. Thus, in collaboration with workers at IITA, we reported the occurrence of EACMV in Nigeria[1], and more recently showed that all three African viruses occur in Madagascar[2]. However, the most interesting work from an evolutionary point of view centred on Uganda.

Cassava is not a native plant to Africa but was introduced from South America, probably by Portuguese traders and settlers. When it first reached Uganda is unclear, but its presence there was noted by Speke during his expedition in 1875. Until the 1920's it was an unimportant crop, but thereafter its cultivation increased and it is now the main staple in much of the country. Cassava mosaic disease, caused by ACMV, was first recorded in 1928 and subsequently became endemic, although the losses it caused were tolerable. This situation changed from 1988 when an outbreak of severe disease was reported in Luwero district and spread southwards at a rate of 20 km / year, eventually reaching the northern shores of Lake Victoria. In the area affected by the epidemic, which has now spread into all neighbouring countries, cassava yields were virtually nil and people starved.

Workers at SCRI contributed to understanding the causes of this epidemic by characterizing a novel form of begomovirus from cassava affected by the severe disease[3]. This virus, called the Uganda variant (UgV), has a DNA-A that is clearly the product of recombination between the DNA-A's of ACMV and EACMV. Most of the sequence is like that of EACMV but the coat protein gene is like that of ACMV (Fig 1). In both parts, the sequences faithfully match those of the parent viruses, and examination of a range of UgV isolates showed that there was hardly any variation among them. These observations implied that the recombination event that led to the formation of UgV was recent, and that UgV had not yet had time to diversify by the accumulation of mutations. UgV occurred only in the area affected by the epidemic, indicating a causal connection between the new virus and the epidemic. Moreover, in a popular Ugandan cassava variety, UgV reached about 20-fold higher concentration than did an isolate of ACMV typical of those circulating in Uganda before the epidemic. However, the most severe disease appeared in plants infected with both UgV and ACMV[4] (Fig. 2).

UgV is therefore an example of a newly evolved virus with increased pathogenic potential. The key event in its evolution was recombination between the DNA-A's of ACMV and EACMV, which can have taken place only in a cassava plant doubly infected with the two parental viruses. Such double infections are uncommon, and therefore recombination should be a rare event. However, once this particular recombinant had been produced, it was well adapted for transmission by the local strains of whitefly, and therefore it persisted and spread.

Figure 2 Mosaic-affected cassava shoots: left, infected with ACMV; right, infected with UgV; centre, doubly infected with both ACMV and UgV.

Cotton leaf curl disease in Pakistan. Another example of a disease caused by begomoviruses is leaf curl disease of cotton (Fig. 3), which was first recorded in Pakistan in 1967. The incidence of the disease increased rapidly from 1989, coinciding with the widespread cultivation of S-12, a high-yielding long-staple cultivar that is highly susceptible to infection with cotton leaf curl virus. Another factor in the epidemic was the presence of huge uncontrollable populations of whiteflies that had arisen following the over-use of pesticides and the development of resistance to many of them. By 1993-94, losses of cotton (Pakistan's principal export) were estimated at about 2 million bales, worth around $400 million.

There was a possibility that there might be parallels with the cassava mosaic epidemic in Uganda, but when we examined the DNA-A of viruses in leaf curl-affected cotton, it quickly became clear that the situation was rather different. Instead of a single novel form, we found many different variants[5]. Some included regions whose sequence closely resembled that of parts of the DNA-A of okra yellow vein mosaic virus and thus were obvious recombinants, although the recombinants typically induced leaf curl rather than yellow vein symptoms in okra (Fig. 3). The origin of other variants was less clear. The recent evolution of cotton leaf curl virus in Pakistan had therefore been characterized by an explosion of new variants, perhaps induced by the prevalence of susceptible cotton cultivars and massive populations of the vector. In another contrast with the Ugandan cassava scenario, multiple infections were common in Pakistan. More than half the plants tested contained two or more virus variants[6]. There were therefore plenty of opportunities for recombination to produce new forms, but other processes may also have had a role in their evo-

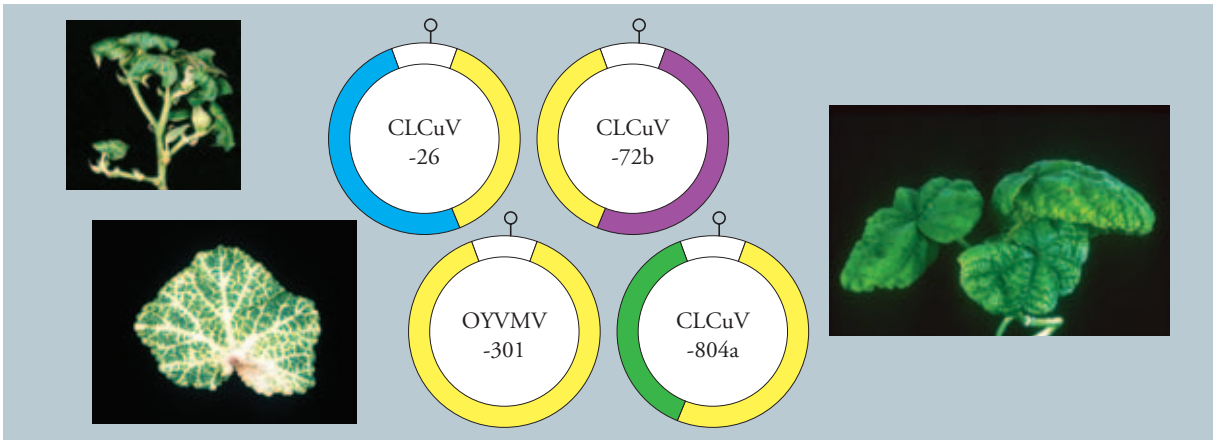

Figure 3 Centre panel: diagrams of the genomes of three recombinant CLCuV variants, showing the regions derived from okra yellow vein mosaic virus (OYVMV) in yellow. Top left: symptoms of CLCuV in cotton. Bottom left: symptoms of OYVMV in okra. Right: symptoms of CLCuV in okra.

lution. As in Uganda, the epidemic was brought under control by the introduction of resistant cultivars. But in Pakistan, the evolutionary pressure was so great that in 2001 a strain of virus appeared that could overcome this resistance, something that has thankfully not yet happened in Uganda.

Evidence of selection pressures

The appearance of new viruses is evidence of the products of evolution, but Darwinian evolution postulates that for new forms to persist they must be well fitted to withstand the selection pressures imposed upon them. For UgV, we can surmise that key factors were its ability to replicate to high concentrations, and its acquisition of the coat protein gene from a local ACMV strain that was presumably already well adapted for transmission by the local whitefly population. In Pakistan it may be that some of the cotton leaf curl virus variants we observed will not survive for long, but that one or two of the best adapted will become dominant. In other cases, the study of patterns of variation among strains of a virus can offer clues to the selection pressures that have operated to bring about the present diversity.

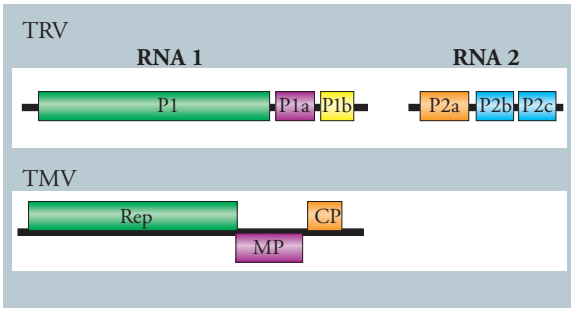

Figure 4 Diagrams of the genomes of tobacco rattle virus (TRV) and tobacco mosaic virus (TMV). Coloured blocks represent genes encoding replication proteins (green), cell-to-cell movement proteins (purple), coat proteins (orange), seed transmission protein (yellow) and nematode transmission proteins (blue).

In contrast to the multiplicity of begomoviruses, the genus *Tobravirus* contains only three species[7]:

Tobacco rattle virus (TRV) occurs in Europe, North America, Japan, China and New Zealand, and infects more than 400 species of monocots and dicots in over 50 families, including many crops and common weeds. It is best known in Britain as one of the causes of spraing disease of potatoes.

Pea early-browning virus (PEBV) occurs in Europe and North Africa, and infects legumes.

Pepper ringspot virus occurs in South America, and causes diseases in peppers, tomatoes and artichokes.

All three viruses are spread by free-living nematodes in the genera *Trichodorus* and *Paratrichodorus* (trichodorids), which are found mainly in light or sandy soils. The genomes of tobraviruses consist of linear single-stranded RNA, and are bipartite. The larger RNA, RNA-1, codes for replication and intercellular movement functions and for a protein involved in seed transmission (Fig. 4). Although there are large differences between the RNA-1 sequences of the three virus species, there is little difference among those of different strains of TRV. In the three TRV strains whose RNA-1 has been completely sequenced, more than 99% of the 6791 bases in RNA-1 are the same in all

three. The smaller RNA, RNA-2, encodes the virus coat protein and other proteins involved in transmission by nematodes (Fig. 4), and there are large differences among the RNA-2 sequences of different TRV strains[7].

The variability of the part of RNA-2 that encodes the virus coat protein is reflected in the extreme serological diversity of virus strains; multiple serotypes of both TRV and PEBV are recognized. We also showed that RNA-2 encodes the determinants of nematode transmissibility[8]. At least 10 trichodorid species can act as vectors, and for TRV there is a substantial degree of specificity between vector species and virus serotype[9]. In some instances, the specificity between a single trichodorid species and a single virus serotype seems absolute[10], but in others the correlation is less exact. This probably reflects the involvement of the other RNA-2-encoded genes in nematode transmission[11]. These observations suggest that the reason why tobravirus RNA-2 is so much more variable in sequence than RNA-1 is that this variation represents adaptation for transmission by different species of nematode vector, or even by different populations of a vector species. In this connection, it should be remembered that, unlike viruses with aerial vectors, soil-borne viruses have limited means for long distance dispersal to new sites. Trichodorids probably move laterally no more than 1 m/yr, and tobraviruses therefore have to rely on means such as soil movement or the movement of infected vegetative propagating material or seeds for their long distance spread. Tobravirus populations are thus effectively evolving in isolation from each other. In these circumstances, properties such as efficient transmission by the local trichodorid population and a wide host range have obvious value for survival.

The diversity of viruses

From examples such as those discussed above, it is clear that evolution driven by natural selection can account for the present day diversification of viruses and for the diversity found within genera and species. When the scope is widened however, the arguments become more speculative. One might ask, for example, why the tobravirus genome is in two parts. Here, it is instructive to compare the tobravirus genome with that of tobacco mosaic virus (TMV), a virus with perhaps the simplest genome of all, consisting of only a single piece of RNA. The TMV genome encodes only four gene products: two proteins involved in RNA replication, a cell-to-cell movement protein and a coat protein (Fig. 4). Tobraviruses have analogues of each of these gene products, which bear sufficient similarities to the TMV proteins to suggest that they have a common origin. The tobravirus genome also encodes three additional proteins: one involved in transmission through seeds and two involved in transmission by nematodes. There are no clues as to the origins of these additional genes, but it seems likely that a common ancestor of TMV and tobraviruses acquired them and subsequently evolved into the tobraviruses. What was the selection pressure that drove this augmented genome to divide into two parts? One possibility is that with the additional genes it was too large for efficient replication. RNA replication is relatively error prone, and as an RNA gets larger the probability of producing an error-free copy falls off rapidly. Another possibility is that hiving off the genes involved in nematode transmission into a separate piece of RNA facilitated their rapid adaptation to different vector populations.

The functions of RNA replication, cell-to-cell movement and coat protein provided by the TMV genome are the minimum set of functions required by an autonomous RNA plant virus. Coat proteins though are multi-functional. They not only protect virus RNA from degradation, but are usually required for plant-to-plant transmission by vectors and for systemic movement of virus through the vascular system of an infected plant. However, a few plant viruses, notably the umbraviruses, do not have a coat protein gene[12]. Umbraviruses use the coat protein of a helper virus to encapsidate their RNA for plant-to-plant transmission by aphids, and thus can only be transmitted from plants that are doubly infected by the umbravirus and the helper. Umbravirus genomes do contain genes encoding RNA replication and cell-to-cell movement proteins, and also include a gene known as ORF3 (Fig. 5), whose protein product provides some of the other functions of a coat protein. Thus the ORF3 protein forms nucleoprotein particles with the virus RNA[13] (Fig. 5), protects the RNA from degradation, and mediates its transport through the phloem[14]. One might speculate that the ORF3 protein is a proto-coat protein in the course of evolution, which has yet to develop the ability to form conventional virus particles and mediate vector transmission. Alternatively, it could be a relic of a fully functional coat protein that has lost its particle formation and vector transmission functions after the virus acquired the ability to use a helper virus protein for these purposes. Of course, neither of these ideas might be correct, and they may merely reflect the prejudice of a

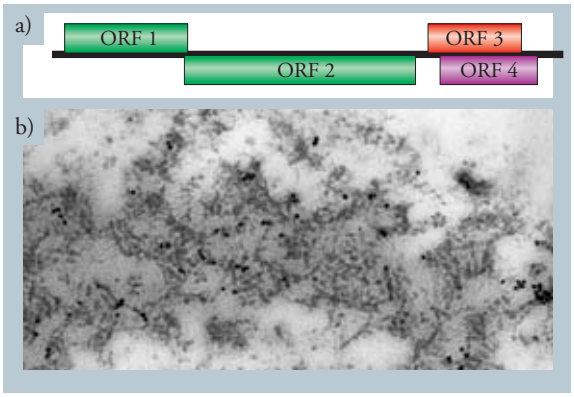

Figure 5 a) Diagram of the genome of an umbravirus. Coloured blocks represent genes encoding replication proteins (green), cell-to-cell movement protein (purple) and ORF3 protein (red). b) Electron micrograph of nucleoprotein filaments containing ORF3 protein and viral RNA in the cytoplasm of an infected cell.

virologist that it is "normal" for a virus to have a coat protein. Moreover, it is difficult to see how they could be tested experimentally.

Although this kind of speculation about the origins of the many types of virus that now exist is an interesting intellectual exercise, it is of little relevance to practical problems of virus pathology. There may be no conclusive answer to the question whether or not a virus is an organism; indeed, as Professor Joad might have said, it all depends what you mean by an organism. But for many purposes, they can be treated as if they were organisms. In particular, the appearance of new variants and the forces that drive these changes seem to follow the same evolutionary models as apply to conventional organisms. If they were pressed on more philosophical questions though, most virologist would simply go along with Lwoff's dictum that "Viruses are viruses".

References

[1] Ogbe F O; Atiri G I; Robinson D; Winter S; Dixon A G O; Quin F M; Thottappilly G (1999). First report of East African cassava mosaic begomovirus in Nigeria. *Plant Disease* **83**, 398.

[2] Ranomenjanahary S; Rabindran R; Robinson D J (2002). Occurrence of three distinct begomoviruses in cassava in Madagascar. *Annals of Applied Biology,* **140**, 315-318.

[3] Zhou X; Liu Y; Calvert L; Munoz C; Otim-Nape G W; Robinson D J; Harrison B D (1997). Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *Journal of General Virology* **78**, 2101-2111.

[4] Harrison B D; Zhou X; Otim-Nape G W; Liu Y; Robinson D J (1997). Role of a novel type of double infection in the geminivirus-induced epidemic of severe cassava mosaic in Uganda. *Annals of Applied Biology* **131**, 437-448.

[5] Zhou X; Liu Y; Robinson D J; Harrison B D (1998). Four DNA-A variants among Pakistani isolates of cotton leaf curl virus and their affinities to DNA-A of geminivirus isolates from okra. *Journal of General Virology* **79**, 915-923.

[6] Sanz A I; Fraile A; García-Arenal F; Zhou X; Robinson D J; Khalid S; Butt T; Harrison B D (2000). Multiple infection, recombination and genome relationships among begomovirus isolates found in cotton and other plants in Pakistan. *Journal of General Virology* **81**, 1839-1849.

[7] Robinson D J & Harrison BD (1985). Unequal variation in the two genome parts of tobraviruses and evidence for the existence of three separate viruses. *Journal of General Virology* **66**, 171-176.

[8] Ploeg A T; Robinson D J; Brown D J F (1993). RNA-2 of tobacco rattle virus encodes the determinants of transmissibility by trichodorid nematodes. *Journal of General Virology* **74**, 1463-1466.

[9] Ploeg A T; Brown D J F; Robinson D J (1992). The association between species of *Trichodorus* and *Paratrichodorus* vector nematodes and serotypes of tobacco rattle tobravirus. *Annals of Applied Biology* **121**, 619-630.

[10] Ploeg A T; Brown D J F; Robinson D J (1992). Acquisition and subsequent transmission of tobacco rattle virus isolates by Paratrichodorus and Trichodorus nematode species. *Netherlands Journal of Plant Pathology* **98**, 291-300.

[11] MacFarlane S A (1999). Molecular biology of the tobraviruses. *Journal of General Virology* **80**, 2799-2807.

[12] Taliansky M E; Robinson D J; Murant A F (1996). Complete nucleotide sequence and organization of the RNA genome of groundnut rosette umbravirus. *Journal of General Virology* 77, 2335-2345.

[13] Taliansky, M; Roberts I M; Kalinina N; Ryabov E V; Raj S K; Robinson D J; Oparka K J (2003). An umbraviral protein, involved in long-distance RNA movement, binds viral RNA and forms unique, protective ribonucleoprotein complexes. *Journal of Virology* 77, 3031-3040.

[14] Ryabov E V; Robinson D J; Taliansky M (2001). Umbravirus-encoded proteins both stabilize heterologous viral RNA and mediate its systemic movement in some plant species. *Virology* **288**, 391-400.