Evolution: **The long evolutionary reach of viruses**

Roger W. Hendrix

The structure of a phage capsid protein provides good evidence this phage shares ancestry with an animal virus. In this and similar cases, either the viral lineages predate the emergence of the three contemporary domains of life, or viruses have been leaping the phylogenetic chasms that separate the domains.

Address: Pittsburgh Bacteriophage Institute and Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA.
E-mail: rhx@vms.cis.pitt.edu

Current Biology 1999, 9:R914–R917

0960-9822/99/$ – see front matter

We are well into the Age of Genomics, and genome sequences are providing a flood of valuable new information about (among other things) the evolutionary histories and relationships of contemporary organisms. This is nowhere more true than among the viruses, where the economical sizes of the genomes mean that large numbers of genomic sequences can be determined and compared; as a consequence, there are now large databases cataloging the sequence relationships and inferred phylogenetic relationships among such notable groups as the Herpesviruses or strains of HIV.

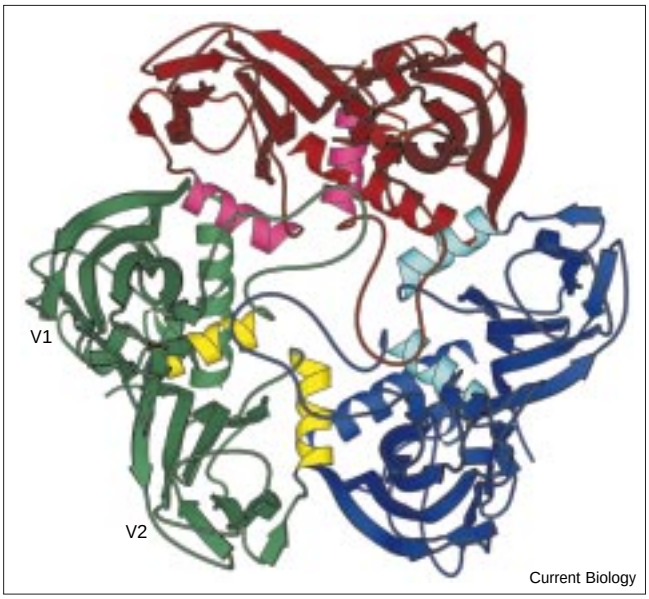
Viral genome sequences are also remarkably diverse, however, and this places severe constraints on the lengths of the evolutionary distances that can be detected by sequence comparisons. For example, among the double-stranded DNA tailed bacteriophages, any two genome sequences picked at random from the population are unlikely to have more than fleeting sequence similarity — this despite the fact that there are many good reasons to believe that this group of viruses shares common ancestry [1]. Apparently the sequences of their DNA and of their proteins have diverged past the point of recognition, even though features of their lifestyles are strikingly conserved. Thus, while a comparison of the genomes of any two animals, say *Drosophila* and humans, would reveal many genes with clear homologs in both species, bacteriophages as a group are not so similar to one another.

If sequences are so different among this one group of viruses, what is the hope of detecting evidence for evolutionary relationships — if they exist — over even greater distances? Happily, it appears that compelling evidence for shared ancestry can survive in viral phenotypes, long after any hint of sequence similarity has departed, and this sort of similarity can be detected spanning even the divide between animal viruses and bacteriophages. The latest example of this comes from structural studies on virus capsid proteins, described in a recent paper from the laboratories of Roger Burnett and Dennis Bamford [2].

Using X-ray crystallography, Benson *et al.* [2] have determined at high resolution the structure of the main capsid subunit, P3, of the double-stranded DNA phage PRD1, which infects *Escherichia coli* and other enteric hosts. The 1.85 Å resolution structure of P3, elegant though it is in isolation, gains its full evolutionary fragrance in comparisons with other virus capsid protein structures — first, with all the other solved structures, and then more specifically with the 'hexon' capsid subunit of Adenovirus, which was determined several years ago in Burnett's lab [3].

Of the roughly 50 virus capsid protein structures that are known to high resolution (see http://mmtsb.scripps.edu/viper.html), most contain a characteristic protein fold known as a 'viral β-barrel' or 'viral jelly roll'. This is an eight-stranded antiparallel β structure with a particular

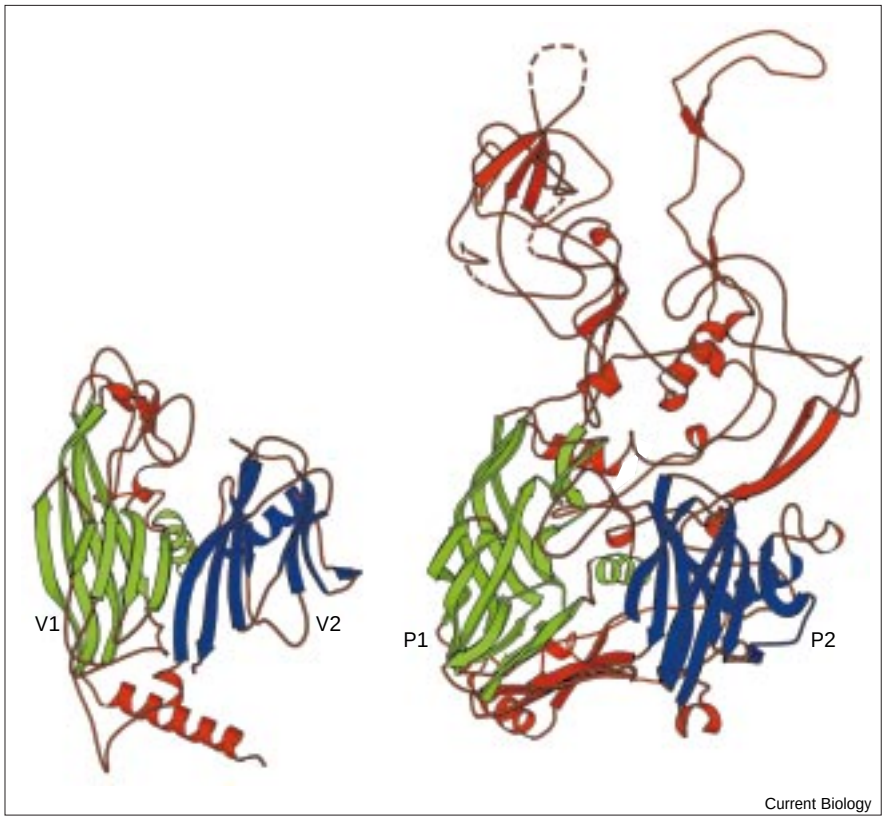
Figure 1



The P3 protein trimer of phage PRD1, viewed from outside the capsid looking down on the surface. Each of the three subunits has two domains, each consisting primarily of a jelly roll fold, and the six jelly roll domains are arranged as seen here to give the sixfold quasi-symmetry of the hexon. The two jelly roll folds of one of the subunits are identified as V1 and V2. (Adapted with permission from [2].)

Figure 2

Comparison of the PRD1 P3 monomer (left) and the Adenovirus hexon monomer (right). The two jelly rolls in each subunit are indicated with green and blue, and labeled as V1 and V2 for the PRD1 protein and as P1 and P2 for the Adenovirus protein. (Adapted with permission from [2].)



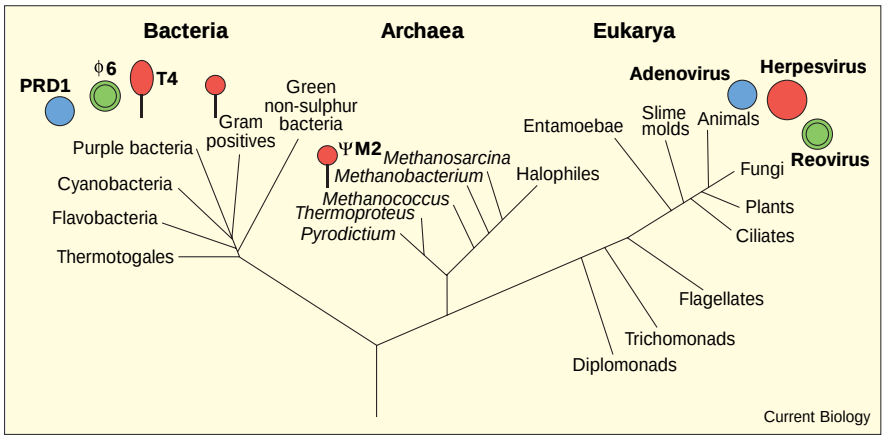
topology of connectivity among the strands. The viruses that share this fold include ones that infect plants (TBSV, TYMV, others), animals (SV40, Rhinovirus, Poliovirus, others) and bacteria (bacteriophage ϕX174). This is, however, not the only successful way to construct an icosahedral virus capsid, as there are viruses — including ones such as Sindbis virus with eukaryotic hosts, and ones such as phage MS2 with prokaryotic hosts — that get by perfectly well with capsid proteins that have completely different folds.

These facts provide suggestive, if not entirely compelling, evidence that the viruses with capsid subunits that have the jelly roll fold are members of a family related by descent from a common ancestor — even though the amino-acid sequences of their capsid proteins have by now diverged past recognition — rather than examples of spectacular evolutionary convergence on the 'optimal' design for a capsid protein fold. Bacteriophage PRD1 and the animal virus Adenovirus both have capsid proteins with jelly roll folds, and so by that criterion are presumptively related.

The particularity of the similarities between the PRD1 and Adenovirus proteins, however, make a much stronger case for common ancestry than their sharing of jelly rolls alone. The Adenovirus hexon protein is so-called because it assembles into an oligomer that occupies the positions of local sixfold symmetry in the capsid. In most viruses, this capsid position is occupied by a hexamer of the subunit protein, but Adenovirus makes its hexon with a trimer of subunits. It accomplishes this by the stratagem that each subunit has two domains, each with its own viral jelly roll fold and each of which occupies the position of a single subunit in a more conventional virus. The newly determined structure [2] shows that the same description applies to the P3 protein of phage PRD1, as shown in Figure 1. Other similarities between the structures of the two viruses' hexon proteins (Figure 2) include the specific topological connectivity within and between the jelly rolls, and the orientation of the jelly roll axes roughly normal to the capsid surface (which is different from other viruses).

The most obvious difference between the two proteins is that the Adenovirus hexon structure has an elaboration of loops (the 'tower') above the jelly roll domains — that is, on the outside of the capsid — that are largely missing in the PRD1 structure. Benson *et al.* [2] suggest that these loops are the part of the protein that interacts with the host's immune system and therefore not needed by the bacteriophage. Support for this view comes from the observation that the highly variable regions of the Adenovirus hexon sequence map to the tower [4].

Figure 3



The long evolutionary reach of viruses. The colored symbols represent the viruses discussed in the text; they are placed next to their hosts on the universal tree of life. Shared colors indicate the evolutionary connections inferred from the similarities discussed in the text.

There are, in addition, several other common features of the two viruses that independently argue for an evolutionary connection. These include their capsid geometry (T = 25, not found in any other known virus) with the pentamer positions occupied by a separate protein; the presence of spikes emanating from the center of the pentons with a role in host recognition and DNA delivery; and inverted terminal repeats on the double-stranded DNA genome, with a covalently bound protein attached to the 5′ ends of the strands that functions in priming DNA replication [5,6]. Although it is not clear how to assign numerical probabilities that any of these similarities, individually or jointly, have occurred by chance, the conclusion that these two viruses share common ancestry seems nonetheless inescapable.

Is this apparent evolutionary connection between Adenovirus and phage PRD1 an aberration, or are there other cross-domain affinities between viruses? Students of virus assembly will know that the Herpesviruses are remarkably like the tailed phages in the assembly and structure of their virions. The following description of capsid assembly applies equally well to the Herpesvirus HSV-1 and to any of a number of tailed bacteriophages (coliphage T4, for example). An icosahedrally symmetric procapsid is assembled, in which a shell of capsid subunits surrounds an internal core of ‘scaffolding protein’. The scaffolding protein is a highly asymmetric and highly α-helical molecule that, following completion of assembly, is cut into fragments by a virally encoded protease and expelled. Replicated DNA is packaged into the resulting empty shell, reaching a final packing density of approximately 40 base pairs per 10^5 cubic Ångstroms — about the same as crystalline DNA and substantially higher than in other classes of viruses [7].

Concomitant with DNA packaging, the icosahedral capsid undergoes a dramatic conformational shift, in which the overall shape changes from nearly spherical to angular and icosahedron-like, the surface changes from highly convoluted to relatively smooth, the hexons change from distorted to symmetrical, and the stability of the structure increases dramatically. This conformational change makes available binding sites for accessory capsid proteins, which bind and further stabilize the capsid [8,9].

Nearly all of this detailed similarity between the molecular behavior of coliphage T4 and HSV-1 capsid proteins takes place in the absence of sequence similarity between the proteins. (Actually, there is a hint of sequence similarity for the DNA packaging terminase proteins, and the functional order of capsid and capsid assembly genes is similar.) As in the case of Adenovirus and PRD1, the HSV-1 versus T4 comparison suggests strongly that these viruses have a common ancestry.

The third pair of eukaryotic and prokaryotic viruses that show convincing similarity consists of the Reoviruses, which infect both plants and animals, and a family of bacteriophages typified by *Pseudomonas* phage ϕ6. Both of these groups have segmented, double-stranded RNA genomes packaged in an unusual double-shelled capsid. The inner shell is a structurally bizarre, 120 subunit $T = 1$ structure, and the outer shell is a more conventional $T = 13$ (780 subunit) shell. At each corner of the icosahedron there is a polymerase and packaging complex that is responsible for both replication and transcription of the genome, and for accurately packaging exactly one of each genome segment into each progeny virion [10,11]. Once again, it is much more plausible to suppose that these detailed similarities are the consequence of common ancestry than convergent evolution.

Finally, there is also a group of viruses that straddles the phylogenetic chasm between the two domains of prokaryotes — the Bacteria and the Archaea. Among the viruses found to infect the Archaea, a substantial fraction

have the appearance of tailed bacteriophages, including examples with both contractile and non-contractile tails. The genome sequence of one of these, ΨM2, was determined recently [12], and a few of its genes were found to show plausible sequence similarity to bacteriophages that infect *Bacillus* and other Gram-positive hosts. Thus, by a more conventional criterion, this Archaevirus and tailed bacteriophages — or at least some of their genes — share ancestry, as might already have been suspected on the basis of conservation of their characteristic virion morphology.

If we accept these cross-domain congruences as truly representing evolutionary connections (and I for one find it hard to take the alternatives very seriously), then what is the historical basis of the connections? This is, of course, complete speculation, as among other problems we currently have no means to determine the time of occurrence of any evolutionary event we might wish to postulate. However, there would seem to be two mechanisms to explain the observed properties of the viruses — either parallel descent from a population of viruses that were active prior to the divergence of the three contemporary domains, or horizontal exchange of viruses or parts of viral genomes across domain boundaries (see Figure 3).

For the genes cited above that retain sequence similarity between the viruses of Archaea and of Bacteria, I would favor horizontal transfer of those parts of the genomes as the most likely explanation. But to explain the appearance of phage-like viruses on both sides of the Bacteria/Archaea divide, as well as the examples of Bacteria–Eukaryote connections given above, it seems easier to postulate the existence of ancestral viruses already resembling these contemporary forms, that predated the divergence of life into the current divisions. If so, viruses are (perhaps not surprisingly) truly ancient, and their role in the early evolution of life may well have been more profound than is generally realized [13].

References

1. Hendrix RW, Smith MCM, Burns RN, Ford ME, Hatfull GF: Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proc Natl Acad Sci USA* 1999, 96:2192-2197.
2. Benson SD, Bamford JKH, Bamford DH, Burnett RM: Viral evolution revealed by Bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* 1999, 98:825-833.
3. Roberts MM, White JL, Grütter MG, Burnett RM: Three-dimensional structure of the adenovirus major coat protein hexon. *Science* 1986, 232:1148-1151.
4. Rux JJ, Burnett RM: Type-specific epitope locations revealed by X-ray crystallographic study of Adenovirus Type 5 hexon. *Mol Ther* 1999, in press.
5. Rydman PS, Caldentey J, Butcher SJ, Fuller SD, Rutten T, Bamford DH: Bacteriophage PRD1 contains a labile receptor-binding structure at each vertex. *J Mol Biol* 1999, 291:575-587.
6. Burnett RM: The structure of adenovirus. In *Structural Biology of Viruses.* Edited by Chiu W, Burnett RM, Garcea RL. Oxford: Oxford University Press; 1997:209-238.
7. Casjens SR: Principles of virion structure, function, and assembly. In *Structural Biology of Viruses.* Edited by Chiu W, Burnett RM, Garcea RL. Oxford: Oxford University Press; 1997:3-37.
8. Black LW, Showe MK, Steven AC: Morphogenesis of the T4 head. In *Molecular Biology of Bacteriophage T4*. Edited by Karam JD, Drake JW, Kreuzer KN, Mosig G, Hall DH, Eiserling FA, Black LW, Spicer EK, Kutter E, Carlson K, Miller ES: ASM Press; 1994:218-258.
9. Steven AC, Spear PG: Herpesvirus assembly and envelopment. In *Structural Biology of Viruses.* Edited by Chiu W, Burnett RM, Garcea RL. Oxford: Oxford University Press; 1997:312-351.
10. Butcher SJ, Dokland T, Ojala PM, Bamford DH, Fuller SD: Intermediates in the assembly pathway of the double-stranded RNA virus ϕ6. *EMBO J* 1997, 16:4477-4487.
11. Grimes JM, Burroughs JN, Gouet P, Diprose JM, Malby R., Zientara S, Mertens PP, Stuart DI: The atomic structure of the bluetongue virus core. *Nature* 1998, 395:470-478.
12. Pfister P, Wasserfallen A, Stettler R, Leisinger T: Molecular analysis of ***Methanobacterium*** phage ΨM2. *Mol Microbiol* 1998, 30:233-244.
13. Woese C: The universal ancestor. *Proc Natl Acad Sci USA* 1998, 95:6854-6859.