Arch Virol (2005) 150: 629–635
DOI 10.1007/s00705-004-0466-9

VDN
Virology Division News

Report on an ICTV-sponsored symposium on Virus Evolution

U. Desselberger

Virologie Moléculaire et Structurale, CNRS, Gif-sur-Yvette, France

Received September 29, 2004; accepted November 14, 2004
Published online January 13, 2005

A symposium on Virus Evolution, sponsored by the International Committee on Taxonomy of Viruses (ICTV), was held at the 23rd Annual Meeting of the American Society for Virology (ASV) in Montreal, Canada on July 10, 2004. It was organized by Ann Palmenberg (University of Madison-Wisconsin) and Andrew Ball (President, ICTV; University of Alabama at Birmingham) and was supported by Academic Press/Elsevier, Bristol Myers Squibb, The University of Alabama at Birmingham School of Medicine, US National Biodefense Analysis and Countermeasures Center, Wyeth Lederle Vaccines, and the ASV.

Andrew Ball introduced the symposium by pointing out extremes of the taxonomic tasks demanded by virology, ranging from the ability to distinguish viruses clearly into strains, species, genera, families and orders in some cases (the example of the recently achieved classification of human papillomaviruses was given; de Villiers et al., [1]) to the enormous difficulties caused by extensive mosaicism as encountered in the genomes of dsDNA tailed bacteriophages (e.g. [13]). Such data would require a multidimensional taxonomy for phylogenetically accurate representation. The lectures of the symposium reviewed achievements and addressed problems of viral taxonomy. Such problems are considered to be not just formal systematics but to be profoundly linked to questions of viral replication and evolution.

David Mindell (University of Michigan, Ann Arbor) presented his views on "***Viruses and the tree of life***". Starting from Linnean taxa as a hierarchy of categories and Darwin's concept of the evolution of life (1837), modern concepts of the tree of life have been developed (recently funded by an NSF project to assemble "the tree of life"). After clarification of some basic definitions and terms (life, genes, species etc), possible origins of viruses were considered. The primordial hypothesis assumes that RNA viruses emerged at a very early stage from an (ill-defined) 'origin of life' [18]. By contrast, DNA viruses are generally considered to have evolved from bacteria. However, some DNA viruses of the *Archaea* may also have ancestors that preceded the division into the three domains of life approximately 3 billion years ago [23]. The concept of homology of genes ('same organ in different organisms'; Richard Owen, 1804–72, see [29]) led to the recognition of possible molecular mechanisms of evolution, defined as synology (gene duplication), orthology (exon shuffling), and xenology (lateral gene transfer). A number of putative xenologous genes have been recognized in viruses and non-viruses (bacteria, fungi, eukaryotes) encoding: DNA polymerases, ribonucleotide reductases, thymidine synthetases, oncogenes, receptor

genes etc. However, for viruses it should be noted that not all genes encoding proteins of similar functions (e.g. polymerases) stem from one single root. Viruses in the *Retroviridae* seem to be of very ancient origin as retroelements have been found in *Eukarya*, *Bacteria* and now also in *Archaea*; in the latter, there is evidence for at least four different lateral gene transfers [22]. Regarding the order of genes in a genome, patterns have often been maintained, but also often been rearranged. Phylogenetic relationships have been found useful for virus identification, work on origin, speed and mechanisms of evolution, taxonomy, and the elucidation of transmission pathways (e.g. the transmission of HIV from a source to a victim [16]). A case is being made for the use of rankless taxonomy clades (these being monophyletic groups without grading) instead of hierarchical formal Linnean taxa for classification (see PhyloCode; www.ohiou.edu/phylocode). Such a tree-based, rankless system could be constructed independently of taxonomy.

In the discussion, the positions of clades within accepted phylogenies and the role of the quasispecies concept in a phylogenetically based classification system were considered.

Alexander E Gorbalenya (Leiden University Medical Center) spoke about "***Using evolutionary models to learn about RNA viruses***". Starting from the idea that evolutionary models lead to the generation of structure/function studies, which in turn may or may not verify the model, the concept was developed that *biopolymer alignments represent evolutionary models.* Proof of concept was explored using sequence data for members of the *Flaviviridae*, *Nidovirales* and *Birnaviridae* as examples. Citing data from Lindenbach and Rice [15] and the group of Tautz (e.g. [33]), it was concluded that hepaciviruses (e.g. *Hepatitis C virus*) and pestiviruses (e.g. *Bovine viral diarrhea virus*) have more in common than was originally thought; however, the phylogenetic analysis of hepacivirus and pestivirus genomes is still a challenge to the taxonomy. The *Nidovirales* were established as an order that comprises the *Coronaviridae*, *Arteriviridae* and *Roniviridae* families. This conclusion was based on the finding that viruses in the *Nidovirales* share the mechanism of discontinuous transcription [30] and that they have motifs of their replicase enzymes in common [6, 32]. However, the taxonomy of the *Coronaviridae* is under further review [5]. For instance, it has recently been found that the cysteine proteinases of an invertebrate nidovirus and of members of the *Potyviridae* share unusual motifs [35]. Viruses in the *Birnaviridae* (carrying dsRNA genomes) and some (but not other) members of the *Tetraviridae* (carrying ssRNA genomes and infecting insects) share a unique arrangement of motifs in their replicases [6]. A particular folding model of the replicase of *Infectious bursal disease virus*, a member of the *Birnaviridae*, has recently been tested and verified by the group of E Mundt [34].

In the discussion, the relationship between coronaviruses and influenza C viruses (sharing neuraminate-O-acetyl esterase motifs and functions) was noted.

Graham Hatfull (University of Pittsburgh, Pennsylvania) spoke about "***Mycobacteriophage genomics and the origins of mosaicism***". Given that there are an estimated 10^{31} bacteriophages on earth (most of them in the sea), they represent an enormous genomic diversity and are also an excellent tool box with which to probe evolutionary theories. Approximately 250 tailed dsDNA bacteriophages have been completely sequenced, and 30 of those represent mycobacteriophages (of a genome size of approximately 2 Mbp). Partial genomic sequences of 14 of these phages (approximately 1 Mbp each) have been subjected to phylogenetic comparison and analyses [7, 8]. Their genes are closely packed and code for replication, integration, assembly and regulation functions. In the genomes, there is pervasive mosaicism, implying that horizontal exchange of genes has been an important component of their evolution. Over 80% of the genes are only seen in mycobacteriophages but there are also some non-phage genes (which probably were picked up from host genomes). In terms of evolution and classification, each phage genome is considered to be a unique assembly of individual modules (a module either being an individual gene or a set of genes). In order

to arrive at its present resting place, each module has a different phylogenetic history. The models for the generation of mosaicism are targeted recombination and random illegitimate recombination, followed by selection ('*Recombination reassorts genetic modules*'). In order to conceptualize evolutionary relationships, the model of a three-dimensional web-like (or 'sweb') reconstruction of events was proposed. This would allocate unique 'sequence space' to each phage without ranks or preconceptions.

In the discussion, the issues of the stability of mosaic genomes, the speed of recombination during evolution, the lack of a species concept, and the integration and reactivation of mosaic genomes were considered.

Simon Wain-Hobson (Institut Pasteur, Paris) described and analyzed "***The enormous multiplicity of HIV infection* in vivo *and the end of clonality***". In addition to a minimum point mutation rate of 0.25/genome (increasing to 700/genome when nearing 'error catastrophe', see below), each HIV genome has undergone 3 recombination events on average, i.e. recombination creates much more diversity than point mutations [12, 14, 17]. *In vitro*, a single round of replication of HIV-1 in T lymphocytes generated on average 9 recombination events per virus [14]. HIV recombinants are frequently produced within individuals, and are even more frequently observed at epidemiological levels. SIV recombinants are discovered within 15 days of infection. A prerequisite of recombination is a multiply infected cell (either co- or superinfected); such cells have been found in HIV-infected patients [17]. Proviral sequences are randomly distributed on chromosomes; one chromosome can harbour several of them. There are also recombinant proviruses. There can be 600–700 proviral DNA copies per cell, and the amount of DNA in a cell can be increased threefold. One T cell can produce 500–4000 HIV particles that are spread preferentially by cell-to-cell contact. Within an individual, donor cells (mostly dendritic, i.e. antigen presenting cells) carry sequences different from those found in recipient cells (mostly T cells). Upon multiple infections the recombination rate increases and can reach the level of self-destruction ('error catastrophe', see below). Concomitantly, the ratio [number of virus particles (nvp)/pfu], already high for all members of the *Retroviridae*, increases further. In these circumstances, an accurate phylogeny cannot be constructed.

John Coffin (Tufts University School of Medicine, Boston MA, and National Cancer Institute, Frederick MD) spoke about "***Retrovirus evolution and drug resistance***". For retroviruses, host-virus co-evolution has been known for some time. The formation of endogenous retroviruses (ERVs) as integrated proviral sequences leads to indefinite vertical transmission in the host. ERVs can be considered and analysed as representing fossil records of previous virus-host interactions [9]. In human germlines, HERV-K sequences are ubiquitous [10]. The history of retrovirus evolution in humans is long. ERVs represent 6–8% of the human genome. Strong parallels can be found in the phylogeny of ERV of primates and that of primates themselves to the extent that the time points of evolutionary events in ERVs and primates can be mutually determined. HERV-K sequences entered human hosts approximately 30 million years ago. Every human individual carries 30–50 different ERVs of which 13 are considered as 'old' and 24 as 'new'. The analysis of long terminal repeats (LTRs) of ERVs has allowed distinct waves of infection to be identified. Mutations have accumulated as the species evolved. However, the 5′ and 3′ ends of the LTRs have not always co-evolved (about 6/36 human proviruses have 'mismatched' LTRs). Approximately 50% of sequence changes are consistent with evolution by point mutations; other changes are due to multiple recombination events. HERVs are still active and can be reactivated. Using examples of the development of resistance of HIV to the action of the antiviral drug 3′-thiacytidine (3TC), mutations, selection, drift and linkage were recognized as genetic factors affecting the evolution of drug resistance. By using an ultrasensitive detection assay [20], direct sequencing of HIV RNA from limiting dilutions and the application of mathematical methods, extensive recombination events and evidence for

compensatory mutations were recognized as the main factors in the development of drug resistance.

Esteban Domingo (Centro de Investigación en Sanidad Animal and Universidad Autónoma de Madrid) spoke about "***Quasispecies dynamics and extinction of RNA viruses***". After an introduction in which basic genetic terms were defined (mutation and mutation rate, hypermutation, recombination, reassortment, segmentation etc), the quasispecies concept was presented according to which any sample of an RNA virus represents a 'swarm' of closely related mutants. This composition allows the virus to adapt in a flexible way to changing environmental conditions. Parameters of adaptability are: the number of mutations per genome (1–100), the population size (up to 10^{12} infectious particles/host organism), the genomic length (9.5 kb for HIV, 3-30 kb for other RNA viruses, i.e. relatively small for all RNA viruses), and the number of mutations needed to produce a phenotypic change (can be very small). Mutant spectra matter for the quasispecies of many RNA viruses (vesicular stomatitis virus, picornaviruses [poliovirus, foot-and-mouth disease virus (FMDV)], lymphocytic choriomeningitis virus (LCMV), bunyaviruses etc). Hypermutated (pre-extinction) RNA often interferes with the infectivity of clonal RNAs. Assignment of a quasispecies to a phenotype is indeterminate. Quasispecies have both deterministic and stochastic features [21, 27]. Under bottleneck conditions (e.g. plaque-to-plaque passage in cell culture), the quasispecies spectrum will become narrower, and the fitness of the quasispecies to survive will decrease, due to the operation of Muller's ratchet [3]. The fidelity of the transcriptase/replicase will go in parallel with the viability of a quasispecies distribution; with decreasing fidelity of these enzymes, the viral sequences will transgress via an error threshold to become random sequences. For FMDV, a constant rate of 0.25 mutations/genome/plaque transfer has been found during plaque-to-plaque passage. In the presence of a mutagen, viral extinction was frequently observed *in vitro* [2]. Ribavirin, a licensed antiviral drug, was shown to be a mutagen as well. Chronic infection of mice with LCMV was prevented (cured) by treatment of the animals with fluorouracil, a mutagen [28].

In the discussion, the influence of the ratio [nvp/pfu] on viability was considered.

Marilyn Roossinck (Samuel Roberts Noble Foundation, Ardmore OK) asked "***What determines the quasispecies population size? Lessons from plant viruses***". Using examples from the *Tobamovirus* genus and the *Bromoviridae* family, it was shown that the mutation frequency depended on virus host interactions [25, 26]. For *Brome mosaic virus*, the control of diversity was located in RNA segment 2, encoding the polymerase protein, and RNA segment 3, encoding the cell-to-cell movement and coat proteins. Bottleneck conditions limited diversity: of the 15 (silent) mutants in a mixed inoculum, only 7 were found in the 8th leaf and only 5 in the 15th leaf from the site of inoculation. The transmission frequency differed for different mutants. Viruses with large host ranges were found to have large quasispecies 'swarms' (or 'clouds' [31]). For further details see www.noble.org/virus evolution.

Ann Palmenberg spoke about "***RNA structure and comparative picornavirology***". For RNA viruses, every viral base is to be regarded as a compromise forged by the totality of different selective pressures. Those are mainly: protein recognition, mRNA transcription, mRNA translation, protein structure, and RNA structure. Different nucleic acids occur in different structural forms: *in vivo*, DNA is usually in the B form, containing a wide major groove and rising by 3.4 Å/bp. Duplex regions of RNA occur in the A form, rising by 2.6 Å/bp and being more stable than DNAs. The base stacking of RNAs contributes hugely to their stability, and RNA folding is largely driven by base stacking [4]. Evolutionary co-variance of nucleotides is sometimes observed; compensatory changes may stabilize a stem, and such observations may help to confirm an RNA structure. The most stable forms of RNA or DNA contain a minimum of free energy [36, 37]. Using computer programmes developed by Zuker's group, the optimal folding of RNAs of relatively small size (picornaviruses) and

large size (SARS coronavirus) has been accomplished [19, 24]. The question arose: how can one recognize if a calculated fold is real? After randomizing and refolding the RNA sequence of encephalomyocarditis virus (*in silico*), a highly stable form (ΔG of -1720 Kcal/mol) was obtained that was indistinguishable in stability from, and in the fold of, the real RNA. Thus, in reality 'the optimal fold' should be considered a myth; there is no single optimal structure. By mathematical derivation the number of alternative partners with which each base of an RNA can interact (= *P-num*) can be obtained [36] and plotted against the sequence; troughs of the curve indicate regions of few alternative partners. The *P-num* derivative is a '*quantitative measure of the propensity of that base to become involved with the same or alternative pairing partners in a collection of suboptimal folds*' [11, 19]; it is thus a powerful parameter for locating wriggles in an RNA structure. '*To maintain RNA structure*, *evolution selects against better alternatives elsewhere in the genome*'. The internal ribosome entry site (IRES) of picornaviral RNAs is a structural motif with a low *P-num* value. IRES structures are similar to those of tRNAs in that they exhibit no significant sequence similarity, yet fold into virtually identical structures. The picornaviral *cis*-acting replication elements (CREs), which display a CACAAA sequence to 3D polymerases, also have low *P-num* values, and again very different sequences adopt very similar 3-dimensional structures. *Vice versa*, nucleotide sequence similarity does not always conserve RNA structures. The aim of the talk was to show the significance of RNA structural considerations for the evolution of viruses.

At the conclusion of the symposium, *Andy Ball* thanked all speakers and discussants. The symposium was attended by approximately 150 participants.

Comments and suggestions on drafts of this report by Andrew Ball, Jean Cohen, Esteban Domingo, Mary Estes, Denis Fargette, Anne-Lise Haenni, Mike Mayo and Ann Palmenberg are gratefully acknowledged.

References

1. de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H (2004) Classification of papillomaviruses. Virology 324: 17–27
2. Domingo E, Escarmis C, Lazaro E, Manrubio SC (2005) Quasispecies dynamics and RNA virus extinction. Virus Res (in press)
3. Escarmis C, Davila M, Domingo E (1999) Multiple molecular pathways for fitness recovery of an RNA virus debilitated by operation of Muller's ratchet. J Mol Biol 285: 495–505
4. Freier SM, Kierzek R, Jaeger JA, Sugimoto N, Caruthers MH, Neilson T, Turner DH (1986) Improved free-energy parameters for predictions of RNA duplex stability. Proc Natl Acad Sci USA 83: 9373–9377
5. Gonzalez JM, Gomez-Puertas P, Cavanagh D, Gorbalenya AE, Enjuanes L (2003) A comparative sequence analysis to revise the current taxonomy of the family *Coronaviridae*. Arch Virol 148: 2207–2235
6. Gorbalenya AE, Pringle FM, Zeddam JL, Luke BT, Cameron CE, Kalmakoff J, Hanzlik TN, Gordon KH, Ward VK (2002) The palm subdomain-based active site is internally permutated in viral RNA dependent RNA polymerases of an ancient lineage. J Mol Biol 324: 47–62
7. Hendrix RW, Hatfull GF, Smith MC (2003) Bacteriophages with tails: chasing their origins and evolution. Res Microbiol 154: 253–257
8. Hendrix RW (2003) Bacteriophage genomics. Curr Opin Microbiol 6: 506–511
9. Hughes JF, Coffin JM (2002) A novel, endogenous retrovirus-related element in the human genome resembles a DNA transposon: evidence for an evolutionary link? Genomics 80: 453–455
10. Hughes JF, Coffin JM (2004) Human endogenous retrovirus K solo-LTR formation and insertional polymorphism: implications for human and viral evolution. Proc Natl Acad Sci USA 101: 1668–1672

11. Jaeger JA, Turner DH, Zuker M (1989) Improved predictions of secondary structures of RNA. Proc Natl Acad Sci USA 86: 7706–7710
12. Jung A, Maier R, Vartanian JP, Bocharov G, Jung V, Fischer U, Meese E, Wain-Hobson S, Meyerhans A (2002) Multiply infected spleen cells in HIV patients. Nature 418: 144
13. Lawrence JG, Hatfull GF, Hendrix RW (2004) Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. J Bacteriol 184: 4891–4905
14. Levy DN, Aldrovandi GM, Kutsch O, Shaw GM (2004) Dynamics of HIV-1 recombination in its natural target cells. Proc Natl Acad Sci USA 101: 4204–4209
15. Lindenbach BD, Rice CM (2001) *Flaviviridae*: The viruses and their replication. In: Knipe DM, Howley PM (eds), pp 991–1041. Fields virology, 4th edition. Lippincott, Williams & Wilkins, Philadelphia
16. Metzker ML, Mindell DP, Liu XM, Ptak RG, Gibbs RA, Hillis DM (2002) Molecular evidence of HIV-1 transmission in a criminal case. Proc Natl Acad Sci USA 99: 14292–14297
17. Meyerhans A, Jung A, Maier R, Vartanian JP, Bocharov G, Wain-Hobson S (2003) The non-clonal and transitory nature of HIV *in vivo*. Swiss Med Wkly 133: 451–454
18. Mindell DP, Villareal LP (2003) Don't forget about viruses. Science 302: 1677
19. Palmenberg AC, Sgro JY (1997) Topological organization of picornaviral genomes: Statistical prediction of RNA structural signals. Semin Virol 8: 231–241
20. Palmer S, Wiegand AP, Maldarelli F, Bazmi H, Mican JM, Polis M, Dewar RL, Planta A, Liu S, Metcalf CA, Mellors JW, Coffin JM (2003) New real-time reverse transcriptase-initiated PCR assay with single-copy sensitivity for human immunodeficiency virus type 1 RNA in plasma. J Clin Microbiol 41: 4531–4536
21. Quer J, Huerta R, Novella IS, Tsimring L, Domingo E, Holland JJ (1996) Reproducible nonlinear population dynamics and critical points during replicative competitions of RNA virus quasispecies. J Mol Biol 264: 465–471
22. Rest JS, Mindell DP (2003) Retroids in Archaea: phylogeny and lateral origins. Mol Biol Evol 20: 1134–1142
23. Rice G, Tang L, Stedman K, Roberto F, Spuhler J, Gillitzer E, Johnson JE, Douglas T, Young M (2004) The structure of a thermophilic archaeal virus shows a double stranded DNA viral capsid that spans all domains of life. Proc Natl Acad Sci USA 101: 7716–7720
24. Rowe CL, Fleming JD, Nathan MJ, Sgro JY, Palmenberg AC, Baker SC (1997) Generation of coronavirus spike deletion variants by high-frequency recombination at regions of predicted RNA secondary structure. J Virol 71: 6183–6190
25. Roossinck MJ (2002) Evolutionary history of Cucumber mosaic virus deduced by phylogenetic analyses. J Virol 76: 3382–3387
26. Roossinck MJ (2003) Plant RNA virus evolution. Curr Opin Microbiol 6: 406–409
27. Ruiz-Jarabo CM, Miller E, Gomez-Mariano G, Domingo E (2003a) Synchronous loss of quasispecies memory in parallel viral lineages: a deterministic feature of viral quasispecies. J Mol Biol 333: 553–563
28. Ruiz-Jarabo CM, Ly C, Domingo E, de la Torre JC (2003b) Lethal mutagenesis of the prototypic arenavirus lymphocytic choriomeningitis virus (LCMV). Virology 308: 37–47
29. Rupke N (1995) Richard Owen: Victorian naturalist. Yale University Press, New Haven CT
30. Sawicki SG, Sawicki DL (1998) A new model for coronavirus transcription. Adv Exp Med Biol 440: 215–219
31. Schneider WL, Roossinck MJ (2001) Genetic diversity in RNA virus quasispecies is controlled by host-virus interaction. J Virol 75: 6566–6571
32. Snijder EJ, Bredenbeek DJ, Dobbe JC, Thiel V, Ziebuhr J, Poon LL, Guan Y, Rozanov M, Spaan WJ, Gorbalenya AE (2003) Unique and conserved features of genome and proteome of SARS-coronavirus, an early split off from the coronavirus group 2 lineage. J Mol Biol 331: 991–1004

33. Tautz D, Lassig M (2004) Of statistics and genomes. Trends Genet 20: 344–346
34. von Einem UI, Gorbalenya AE, Schirrmeier H, Behrens SE, Letzel T, Mundt E (2004) VP1 of infectious bursal disease virus is an RNA dependent RNA polymerase. J Gen Virol 85: 2221–2229
35. Ziebuhr J, Bayer S, Cowley JA, Gorbalenya AE (2003) The 3C-like proteinase of an invertebrate nidovirus links coronavirus and potyvirus homologs. J Virol 77: 1415–1426
36. Zuker M (1989) On finding all suboptimal foldings of an RNA molecule. Science 244: 48–52
37. Zuker M (1994) Prediction of RNA secondary structure by energy minimization. In: Griffin AM, Griffin HG (eds), pp 267–294. Methods in molecular biology. Humana Press, Clifton NJ

Author's address: Dr. Ulrich Desselberger, Virologie Moléculaire et Structurale, UMR 2472, CNRS, 1 avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France; e-mail: ulrich.desselberger@vms.cnrs-gif.fr

Verleger: Springer-Verlag GmbH, Sachsenplatz 4–6, 1201 Wien, Austria. – Herausgeber: Dr. M. H. V. Van Regenmortel, École Supérieure de Biotechnologie de Strasbourg (ESBS), Parc d'Innovation, Boulevard Sébastian Brandt, 67400 Illkirch, France. – Redaktion: Sachsenplatz 4–6, 1201 Wien, Austria. – Satz und Umbruch: Thomson Press (India) Ltd., Chennai, India. – Offsetdruck: Holzhausen Druck & Medien GmbH, Holzhausenplatz 1, 1140 Wien, Austria. – Herstellungsort: Wien, Austria. – Printed in Austria.