Available online at www.sciencedirect.com
SCIENCE DIRECT
Virus Research 117 (2006) 90–104
Virus Research
www.elsevier.com/locate/virusres

Review

Topics in herpesvirus genomics and evolution

Duncan J. McGeoch *, Frazer J. Rixon, Andrew J. Davison

Medical Research Council Virology Unit, Institute of Virology, University of Glasgow, Church Street, Glasgow G11 5JR, United Kingdom

Available online 21 February 2006

Abstract

Herpesviruses comprise an abundant, widely distributed group of large DNA viruses of humans and other vertebrates, and overall are among the most extensively studied large DNA viruses. Many herpesvirus genome sequences have been determined, and interpreted in terms of gene contents to give detailed views of both ubiquitous and lineage-specific functions. Availability of gene sequences has also enabled evaluations of evolutionary relationships. For herpesviruses of mammals, a robust phylogenetic tree has been constructed, which shows many features characteristic of synchronous development of virus and host lineages over large evolutionary timespans. It has also emerged that three distinct groupings of herpesviruses exist: the first containing viruses with mammals, birds and reptiles as natural hosts; the second containing viruses of amphibians and fish; and the third consisting of a single invertebrate herpesvirus. Within each of the first two groups, the genomes show clear evidence of descent from a common ancestor, but relationships between the three groups are extremely remote. Detailed analyses of capsid structures provide the best evidence for a common origin of the three groups. At a finer level, the structure of the capsid shell protein further suggests an element of common origin between herpesviruses and tailed DNA bacteriophages.

Keywords: *Herpesviridae*; Genome annotation; Phylogeny; Coevolution; Capsid structure

Contents

1. Introduction

The herpesviruses (HVs) are a group of large DNA viruses with a distinctive virion architecture. Historically, from the 1960s to the 1980s, assignment as a HV was made on the basis of virion morphology. As illustrated in Fig. 1, the HV particle has several distinct components. The genomic DNA is densely packed within an icosahedral ($T = 16$) capsid, which has 162 surface capsomeres and is of diameter around 115–130 nm. Of the capsomeres, 150 are primarily composed of six molecules of one protein species, and 11 are pentamers of the same protein. The final pentameric position is occupied by the portal complex. Enclosing the capsid is an amorphous layer, the tegument, composed of several protein species. This in turn is bounded by the outermost element, a lipid bilayer envelope with embedded protein molecules (typically glycosylated), giving an overall diameter of about 200 nm. The whole structure presents a characteristic appearance in negatively-stained or thin-sectioned electron microscopic images, and was used to define membership of the taxonomic family *Herpesviridae*. Biological criteria were then used to make assignments to three subfamilies, the *Alpha-*, *Beta-* and *Gammaherpesvirinae* (Davison et al., 2005a).

* Corresponding author. Tel.: +44 141 330 4645; fax: +44 141 337 2236.
E-mail address: d.mcgeoch@vir.gla.ac.uk (D.J. McGeoch).

0168-1702/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.virusres.2006.01.002

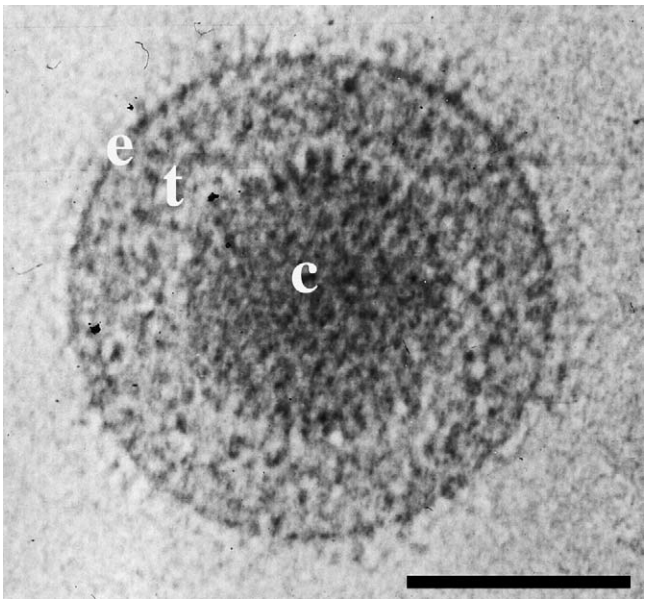

Fig. 1. Electron cryo-microscopic image of an HSV-1 virion. The icosahedral nucleocapsid (c) is embedded in a complex proteinaceous layer called the tegument (t) and enclosed by a lipid envelope containing numerous glycoproteins (e) (bar = 100 nm).

DNA sequence data for HV genomes have been accumulating from the start of the 1980s and are now extensive, with many complete genome sequences and large numbers of partial sequences. It became apparent from sequence-based comparisons that mammalian and avian HVs were descended from a common ancestor and that the three taxonomic subfamilies corresponded to major distinct lineages (with only a few virus species reassigned among subfamilies on the basis of sequence relationships). Sequence comparisons then became the primary approach for evaluating phylogenetic and taxonomic relationships among mammalian and avian HVs, and for identifying newly characterized viruses as members of the *Herpesviridae* and assigning them to subfamilies. Mammalian HVs populate all three subfamilies, while all characterized avian HVs are in the *Alphaherpesvirinae*. Recently it has become clear that characterized reptilian HVs also belong to the *Alphaherpesvirinae* (Herbst et al., 2004; Greenblatt et al., 2005; McGeoch and Gatherer, 2005). A completely different picture has emerged for piscine, amphibian and invertebrate HVs: by criteria of their gene contents, piscine and amphibian HVs form a separate group which appears essentially unrelated to the mammalian/avian/reptilian virus group, and the single known invertebrate HV (of bivalve molluscs) belongs to a third group distinct from both of the vertebrate virus groups (Davison, 2002; Davison et al., 2005b). Thus, at this level, the criterion of a common virion morphology appears to indicate the existence of relationships more distant than can be inferred from DNA or protein sequences.

The *Herpesviridae* Study Group of the International Committee on Taxonomy of Viruses has developed proposals to revise higher level taxonomic arrangements for HVs in order to encompass the above findings. At the time of writing this text, these have not been finally passed into accepted taxonomy. However, with the reasonable expectation that this updated taxonomy will enter general use in the near future, we employ it here. The developments proposed are as follows. First, membership of the family *Herpesviridae* will be restricted to viruses which belong to the *Alpha-*, *Beta-* and *Gammaherpesvirinae* subfamilies on the basis of their gene contents and sequences; these comprise all characterized HVs with mammalian, avian and reptilian hosts. Next, the group of piscine and amphibian HVs will be assigned to a new family, the *Alloherpesviridae*, and the single known invertebrate HV to another new family, the *Malacoherpesviridae*. Third, all three families will be grouped in a new higher level taxon, the order *Herpesvirales*. Additionally, three new genera in the *Herpesviridae* are proposed. *Proboscivirus* in the *Betaherpesvirinae* will contain elephant endothelial HV (EEHV). In the *Gammaherpesvirinae*, the lineage containing alcelaphine HV 1 (AHV-1) and certain other artiodactyl HVs will form the genus *Macavirus*, and the lineage containing equine HV 2 (EHV-2) and certain other perissodactyl and carnivore HVs will form the genus *Percavirus*. Species in both of these last two were formerly assigned to the genus *Rhadinovirus*, which will become more tightly demarcated. These revised arrangements for genera are clarified by the phylogenetic trees in Fig. 4, as discussed in Section 4.

In this review, we first present an evaluation of the current state of HV genomics, and we then describe what can be inferred of the phylogeny of the *Herpesvirales* from analysis of encoded amino acid sequences. Given that 90% of characterized HVs belong to the revised *Herpesviridae*, and also by far the greatest part of experimental investigations have concerned members of this family, our treatment inevitably has these viruses in the foreground. Lastly, we move beyond sequence-oriented analyses to outline deep evolutionary connections that are currently emerging from protein structural investigations, first in the common capsid morphology of the three families of the *Herpesvirales*, and then in similarities that have been observed in the three-dimensional (3D) structures of capsid proteins of HVs and large DNA bacteriophages which may suggest an element of common ancestry for these disparate virus groupings.

2. Herpesvirus genome sequences and their interpretation

Genomes isolated from herpesvirions consist of linear, double-stranded DNA with unpaired, complementary nucleotides at each terminus. Genomes of the *Herpesviridae* range in size from 124 kbp (simian varicella virus (SVV) from the *Alphaherpesvirinae*; Gray et al., 2001) to 241 kbp (chimpanzee cytomegalovirus from the *Betaherpesvirinae*; Davison et al., 2003). The genome of channel catfish virus (CCV), the sole sequenced member of the *Alloherpesviridae*, is 134 kbp in size (Davison, 1992), and viruses in this family that infect carp have the largest known genomes among the *Herpesvirales* (295 kbp; Hutoran et al., 2005; Waltzek et al., 2005). The single member of the *Malacoherpesviridae* (ostreid HV 1 (OsHV-1)) has a genome of 207 kbp (Davison et al., 2005b). Nucleotide compositions range widely, from 32 to 75% G + C, even for viruses in the same genus (Honess, 1984).

HV DNA molecules characteristically contain regions of unique sequence flanked by direct or inverted repeats. Certain