Available online at www.sciencedirect.com
SCIENCE DIRECT
Virus Research 117 (2006) 38–51
Virus Research
www.elsevier.com/locate/virusres

Comparative and functional genomics of closteroviruses

Valerian V. Dolja [a,*], Jan F. Kreuze [b,c], Jari P.T. Valkonen [c,d]

[a] *Department of Botany and Plant Pathology and Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR 97331, USA*
[b] *Germplasm Enhancement and Crop Improvement Division, International Potato Center (CIP), Apartado 1558, Lima12, Peru*
[c] *Department of Plant Biology and Forest Genetics, Swedish University of Agricultural Sciences, SE-75007 Uppsala, Sweden*
[d] *Department of Applied Biology, University of Helsinki, FIN-00014, Finland*

Abstract

The largest extant RNA genomes are found in two diverse families of positive-strand RNA viruses, the animal *Coronaviridae* and the plant *Closteroviridae*. Comparative analysis of the viruses from the latter family reveals three levels of gene conservation. The most conserved gene module defines RNA replication and is shared with plant and animal viruses in the alphavirus-like superfamily. A module of five genes that function in particle assembly and transport is a hallmark of the family *Closteroviridae* and was likely present in the ancestor of all three closterovirus genera. This module includes a homologue of Hsp70 molecular chaperones and three diverged copies of the capsid protein gene. The remaining genes show dramatic variation in their numbers, functions, and origins among closteroviruses within and between the genera. Proteins encoded by these genes include suppressors of RNA silencing, RNAse III, papain-like proteases, the AlkB domain implicated in RNA repair, Zn-ribbon-containing protein, and a variety of proteins with no detectable homologues in the current databases. The evolutionary processes that have shaped the complex and fluid genomes of the large RNA viruses might be similar to those that have been involved in evolution of genomic complexity in other divisions of life.

Keywords: Virus evolution; *Closteroviridae*; Closterovirus; Crinivirus; Ampelovirus

1. Introduction

With their 20–30 kb genomes, the largest known RNA viruses are no match for the champions of the DNA virus world whose genomes reach 1.2 Mb (Iyer et al., 2006), let alone cellular life forms. Nevertheless, RNA viruses play a prominent role in our understanding of life's origin and evolution. Because RNA is widely believed to predate DNA as the genetic material, RNA viruses could be living fossils of the primordial RNA world (Joyce, 2002; Koonin and Martin, 2005). Therefore, at least some genes of RNA viruses are likely to encode extremely ancient, perhaps, primeval proteins involved in replication and metabolism of nucleic acids. From a somewhat different, comparative-genomic perspective, large RNA viruses provide an opportunity to investigate problems of genome complexity and its evolution on a relatively modest, tractable scale.

Closteroviruses share a conserved core of genes involved in replication with other animal and plant viruses within the alphavirus-like superfamily of positive-strand RNA viruses (Dolja et al., 1994). However, closteroviruses stand alone in regard to their genetic capacity and variability, as well as the unique morphology of their particles. It is instructive to compare the genera *Closterovirus* and *Tobamovirus* that both belong to the alphavirus-like superfamily and share the helical virion architecture (Fig. 1). All tobamoviruses have ∼6.5 kb genomes that code for four proteins, one of which assembles to form the rod-shaped virion (Fig. 1B and C). In contrast, the size of closterovirus genomes varies from ∼15.5 to ∼19.5 kb with a coding capacity of 10–14 proteins (Figs. 1A, B and 2). Filamentous virions of closteroviruses incorporate at least five proteins that are assembled into a long body of uniform morphology and a short segmented tail (Fig. 1C). Thus, the larger amount of genetic material in closteroviruses translates into the increased structural complexity and genetic variation among individual viruses.

Here we explore the gene repertoire of the family *Closteroviridae* in an attempt to reconstruct the evolutionary history of these viruses. We also use the available information on the gene functions and regulation to propose a working model of the infection cycle for a 'generic' closterovirus.

* Corresponding author. Tel.: +1 541 737 5472; fax: +1 541 737 3573.
E-mail address: doljav@science.oregonstate.edu (V.V. Dolja).

0168-1702/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.virusres.2006.02.002

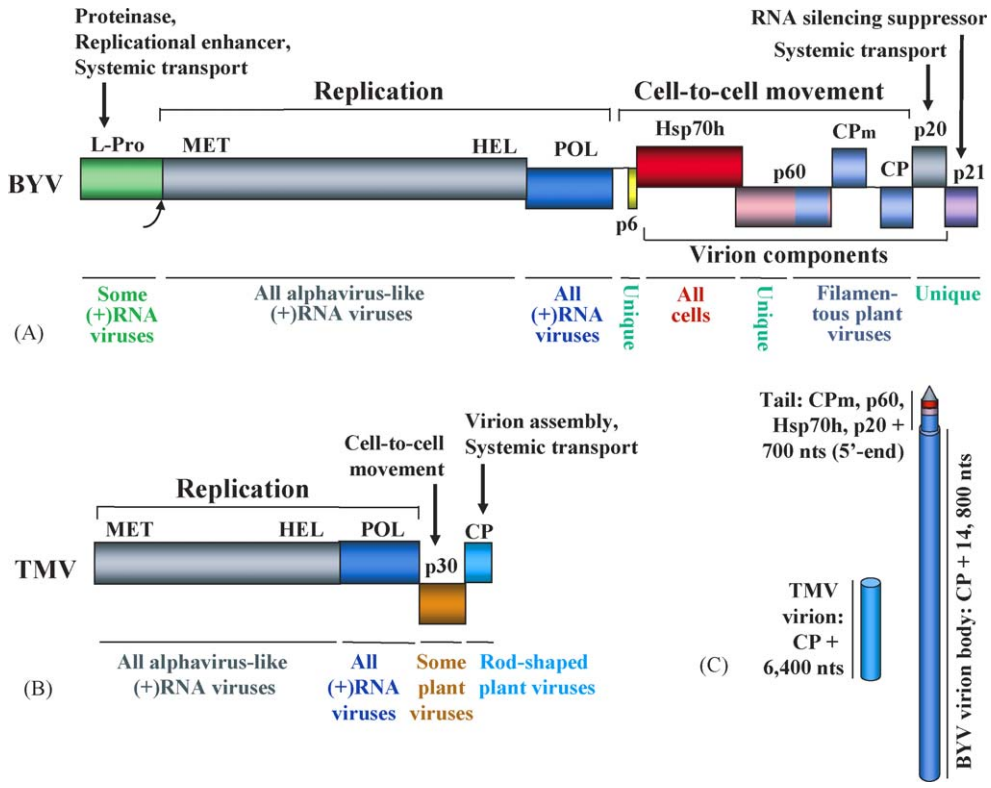

Fig. 1. Comparison of the genetic and structural complexity of *Beet yellows virus* (BYV), genus *Closterovirus*, and *Tobacco mosaic virus* (TMV), genus *Tobamovirus*. (A) Genome map, functions, and evolutionary connections of BYV. The ORFs are shown as cylinders with associated protein designations. L-Pro, leader proteinase; MET, HEL, and POL, methyltransferase, RNA helicase, and RNA-dependent RNA polymerase domains of the replicase, respectively; p6, a 6-kDa protein; Hsp70h, a Hsp70-homologue; p64, a 64-kDa protein; CPm and CP, the minor and major capsid proteins, respectively; p20 and p21, the 20 and 21-kDa proteins, respectively. The sequence similarity between CP, CPm, and the C-terminal domain of p64 is illustrated with the same color. The protein functions are indicated above and below the diagram, while the evolutionary conservation patterns are shown at the bottom. (B) Genome map, functions, and evolutionary connections of TMV. P30, a 30-kDa protein. Other designations are the same as in (A). (C) Cartoons of the TMV and BYV virions. The TMV virions are rigid helical rods of ~300 nm in length and 18 nm in diameter. The BYV virions are flexuous, helical filaments of ~1400 nm in length and 12 nm in diameter with the ~100 nm-long tails that are ~8 nm in diameter. The protein composition and the length of the encapsidated viral RNA is shown.

2. Replication-related genes

As is the case for all positive-strand RNA viruses, putative components of the closteroviral RNA replicase are expressed directly from the virion RNA (Karasev et al., 1989). The product of the 5′-terminal open reading frame (ORF) contains RNA methyltransferase (MET) and RNA helicase (HEL) domains. The RNA-dependent RNA polymerase (POL) is encoded by a downstream ORF that is presumably expressed via a +1 translational frameshift (Agranovsky et al., 1994; Karasev et al., 1995). Thus, translation of the genomic RNA results in two large polyproteins, one spanning MET–HEL, and the other one encompassing MET–HEL–POL; this second, larger polyprotein is produced in much smaller quantities due to the low frequency of frameshifting. The MET–HEL–POL module of closteroviruses is universally conserved in the entire alphavirus-like superfamily (Fig. 1A and B) (Koonin and Dolja, 1993).

A peculiar feature of the closteroviral replicases is the presence of a large variable region between the MET and HEL domains. This region is processed by an unknown mechanism to produce separate MET- and HEL-containing products (Erokhina et al., 2000). Because ectopic expression of the MET–HEL region in plants is sufficient to generate these separate products, the processing could involve either a cryptic proteolytic activity within the MET–HEL polyprotein itself, or a cellular protease (Peremyslov and Dolja, unpublished data).

In addition to their enzymatic functions, RNA replicases of alphavirus-like viruses direct assembly of the replication compartments. The principal component of these compartments is a protein shell formed by the MET–HEL protein subunits (Schwartz et al., 2002). This shell is enveloped by a membrane derived from the endoplasmic reticulum (ER) or other endomembrane. In accord with this paradigm, we found that closterovirus infection results in a drastic remodeling of the ER network that is transformed into a large vesicular factory of viral RNA. We also found that transient expression of the MET–HEL polyprotein alone is sufficient to restructure ER (Prokhnevsky et al., unpublished data). As proposed by Schwartz et al. (2002), one or a few POL-containing replication complexes are present within each virus-induced vesicle to amplify the viral genome and to produce subgenomic messenger RNA (sgRNAs).