

# Genome Replication/Expression Strategies of Positive-Strand RNA Viruses: A Simple Version of a Combinatorial Classification and Prediction of New Strategies

EUGENE V. KOONIN

*Institute of Microbiology, USSR Academy of Sciences, 7 Prospekt 60let Oktyabrya, 117811 Moscow, USSR*

Received May 31, 1990

Accepted October 13, 1990

Requests for reprints should be addressed to Eugene V. Koonin, Institute of Microbiology, USSR Academy of Sciences, 7 Prospekt, 60let Oktyabrya, 117811 Moscow, USSR.

Key words: replication, expression, positive strand RNA, genome strategy, RNA virus, genome-linked virus protein

## ***Abstract***

A combinatorial approach to the classification of replication/expression strategies of positive-strand RNA virus genomes is suggested. Eighteen genome strategies defined as combinations of distinct modes of expression and replication are briefly characterized, 10 of which have been actually found in diverse virus groups. The chances for realization of the remaining eight strategies are evaluated. It is demonstrated that positive-strand RNA virus genome strategies are not necessarily monophyletic characters and could, in some cases, evolve convergently.

## **Introduction**

Positive-strand RNA viruses, i.e., viruses whose genome RNA functions also as the mRNA directing synthesis of at least a subset of virus proteins, including the RNA-dependent RNA polymerase, constitute the largest of virus classes. This class accommodates about half of all virus groups presently approved by the International Committee for the Taxonomy of Viruses, i.e., ca. 35 of 70 (1,2). An important descriptor of each of these groups is what is often called *genome strat-*

*egy*, i.e., the repertoire of molecular mechanisms utilized by the virus to express and replicate its genome (e.g., 3,4). The diversity of specific variants of positive-strand RNA virus genome strategies is enormous, almost frustrating, at least upon a superficial glance. Hence, there is a strong need for a rational classification of these strategies.

The greatest achievement of molecular virology over the last several years has been the sequencing of a number of virus genomes. This is particularly relevant for positive-strand RNA viruses, where representative complete genome sequences of over 20 groups have been reported (5,6). Knowing the genome sequence of a virus does not directly disclose its replication and expression mechanisms. Nevertheless, the sequence information is of great value for understanding the genome strategy, allowing verification and correction of the notions based on biochemical evidence. With these data, it now seems timely to discuss a simple version of classification of positive-strand RNA virus genome strategies. The purposes of this classification are multiple: a) methodological and didactical, facilitating conceptualization of specific replication and expression mechanisms; b) heuristic, allowing the prediction of new strategies; c) finally, it is of major interest to assess, for each genome strategy, the cases for its monophyletic or polyphyletic origin, i.e., the relationship between the genome strategies and virus evolution, or in other words, between divergence and convergence in the evolution of genome strategies. In what follows no attempt has been made to present a comprehensive survey of the relevant literature (a hardly feasible task, in fact), and only reviews and selected original papers containing the most illuminating (from the author's point of view) observations are cited.

The basic approach will be the analysis of combinations of distinct modes of replication and expression, and the identification of such combinations in different virus groups (only nondefective, replication-competent viruses will be dealt with, as virus satellites may have other, highly specific lifestyles). For the proposed classification to be useful, it is crucial that the optimal set of characters, i.e., replication and expression modes, be defined. Clearly, it is impossible to include all specific mechanisms. Thus the set of "fundamental" characters is to be delineated somehow. A simple idea is that those mechanisms that are conserved within large groups of viruses, or among several groups, can be considered to be fundamental. This is warranted by the notion that most of the presently recognized virus groups are apparently monophyletic (5,7,8), and the conserved mechanisms are probably ancestral and have been strongly selected for in the course of evolution.

The main problem in positive-strand RNA virus genome expression is the generation of individual viral proteins. The above criterion allow the delineation of the five fundamental modes by which this is achieved:

1. Translation of polycistronic mRNAs with multiple ribosome entry sites yielding mature individual proteins (the prokaryotic mode of translation).

2. Generation of monocistronic subgenomic mRNA.
3. Posttranslational processing of a large primary translation product (polyprotein) mediated (at least partially) by a virus-encoded protease(s).
4. Genome segmentation.
5. The combination of modes 2 and 3 makes up an additional expression pathway, i.e., processing of polyproteins translated from subgenomic mRNAs.

Finally, the combinations of mechanism 4 with each of the other mechanisms are considered. Altogether, this gives nine “permitted” expression strategies (Table 1). It is postulated that combination of mode 1 with modes 2, 3, or 4 is unlikely, based on the following argument. Establishment of such mechanisms as polyprotein processing and subgenomic mRNA formation requires evolutionary breakthroughs, i.e., the development of virus-encoded proteases specific for distinct cleavage sites, and probably of proteins (domains) involved in the initiation of subgenomic mRNA transcription (3). This evolutionary “work” would be redundant for polycistronic RNAs typical of prokaryotic viruses. On the contrary, in eukaryotic viruses, utilization of truly polycistronic mRNAs is precluded by the strong bias against internal initiation by the eukaryotic translation machinery (9); hence, the driving force for evolution of new mechanisms.

Other mechanisms do not fully meet the above criterion. For example, a number of viruses belonging to several groups utilize translational readthrough of stop codons for the controlled generation of distinct proteins (10). It seems, however, that this mechanism is relatively easily gained and lost in evolution, as demonstrated by the fact that, among closely related alphaviruses, some express the nonstructural proteins via readthrough, whereas others lack the respective termination codon (11).

Delineation of distinct traits characterizing virus genome replication is more problematic. Of obvious importance seems to be the existence of two fundamentally different modes of RNA chain initiation, one of which includes utilization of a virus-encoded genome-linked protein (VPg). Though the precise mechanism of VPg action is not understood, one alternative being actual protein priming (12) and the other cleavage of a self-primed terminal hairpin (13), both possibilities are apparently dissimilar from the initiation mechanism in viruses lacking VPg (14). Elongation mechanisms of virus RNA synthesis are poorly studied. Conceivably, upon further analysis new fundamental characters will be revealed, e.g., conservative vs. semiconservative synthesis.

It is the combination of the fundamental characters describing expression and replication of virus genomes that can be reasonably designated *virus genome strategy*. Table 1 presents a classification of such strategies. Each partition of the table corresponds to a distinct genome strategy. Nine fundamental mechanisms of expression can combine with two replication mechanisms, giving a total of 18 distinct genome strategies.

Table 1. Strategies of positive-strand RNA virus genome expression and replication: A combinatorial classification scheme

Expression mechanisms		Replication mechanisms		
		Non-VPg-utilizing replication	VPg-utilizing replication	
Polycistronic mRNA	Nonsegmented genome	ssRNA phages (MS2-related)	?	1
	Segmented genome	?	?	2
Polyprotein processing	Nonsegmented genome	Flaviviruses Pestiviruses	Picornaviruses Potyviruses	3
	Segmented genome	?	Comoviruses Nepoviruses	4
Subgenomic RNA formation	Nonsegmented genome	Carmoviruses Tombusviruses  Potexviruses Carlaviruses Closteroviruses (ACLSV) Tobamoviruses	Luteoviruses (BYDV)	5
	Segmented genome	Dianthoviruses Tobraviruses Tricornaviruses Hordeiviruses	?	6
Polyprotein processing + subgenomic mRNA formation	Nonsegmented genome	Alphaviruses Rubiviruses Coronaviruses Tymoviruses	Sobemoviruses Luteoviruses (BWYV, PLRV)	7
	Segmented genome	Nodaviruses	?	8
Genome segmentation		?	?	9
		A	B	

The names of virus groups supposed to constitute monophyletic higher taxa (see text) are boxed. ACLSV = apple chlorotic leafspot virus; BWYV = beet western yellow virus, PLRV = potato leafroll virus.

**A Brief Survey of the Strategies Actually Exploited by the Well-Studied Positive-Strand RNA Viruses (the Filled Partitions of the Classification Table)**

Of the 18 genome strategies suggested by the present classification scheme, 10 have been actually found in (relatively) well-studied virus groups. Below, we tackle very briefly each of these strategies and list the evidence for evolutionary relationships between groups of viruses with identical and different genome strategies. How specifically such relationships should be defined is a delicate matter, as recombination, sometimes between remote groups, has apparently made major contributions to positive-strand RNA virus evolution (4,15). As a rough guide, we accept here that sequence similarities between proteins involved in genome replication and expression, particularly between the RNA polymerases, offer the best estimate of such relationships (5,16,17).

**A1.** The genome strategy of “three-cistronic” RNA phages (leviviruses), the only known group of prokaryotic positive-strand RNA viruses. Comparative analysis of the genome sequences of phages belonging to different subgroups clearly showed that phylogenetically they are all closely related (18).

**A3.** The strategy typical of two families of animal viruses, flaviviruses, and pestiviruses. Comparison of protein sequences and genome organizations of these viruses suggested that they are probably members of a single monophyletic supergroup (19–21).

**A5.** The strategy employed by several groups of plant viruses (Table 1). One evolutionary compact group includes carmo- and tombusviruses (22,23) and another consists of potex-, carla-, and a subdivision of closteroviruses (24–26). However, the relationship between these two groups, and between at least the first of them and the tobamoviruses, is quite remote (8,16).

**A6.** A strategy found in several plant virus groups. Phylogenetically they obviously belong to different supergroups, e.g., dianthoviruses are related to carmoviruses, and tobnaviruses to tobamoviruses and hordeiviruses, as revealed by sequence comparison of the proteins mediating genome replication (5,27,28).

**A7.** The strategy employed by three groups of animal viruses with large (i.e., among RNA viruses) genomes, i.e., corona-, alpha-, and rubiviruses. While in alphaviruses, and with less detail in rubiviruses, polyprotein processing is a well-established phenomenon (29), for coronaviruses the involvement of virus-specific proteases in the generation of nonstructural proteins has been demonstrated only recently, both by direct experiments and by theoretical identification of protease domains (30,31, and H.-J. Lee, C.-K. Shieh, A.E. Gorbalenya, E.V. Koonin, N. La Monica, J. Tuler, A. Bagdzahdzian, and M.M.C. Lai, *Virology*, submitted). All three groups of animal viruses exploiting this strategy appear to be only

distantly related to each other in evolutionary terms (31,32). Tymoviruses are close relatives of potex-, carla-, and closteroviruses exploiting the A5 strategy (26,33), and the reports on the processing of tymovirus proteins seemed quite unexpected. Recent data, however, leave little doubt that such processing indeed occurs and is affected by a virus-encoded protease (34).

**A8.** This strategy is observed in the nodaviruses, an insect virus family, though processing has been definitely shown for only the capsid protein precursor (35).

**A9.** The genome strategy of two groups of VPg-containing viruses, one of which infects animals (picornaviruses) and the other infects plants (potyviruses). Comparison of their replicative protein sequences and genome organizations suggested complex evolutionary relationships (5,16,36,37). While the overall organizations of the replicative gene arrays are nearly identical, and the polymerases and proteases of picorna- and potyviruses group together upon phylogenetic analysis, their putative helicases belong to two different superfamilies. This might be an evidence of recombination within the replicative gene complex.

**B4.** The genome strategy of two definitely related groups of plant viruses. It is important to note that these groups are closer to picornaviruses than potyviruses are (5,16).

**B5.** This strategy is apparent in only one virus group, a subdivision of luteoviruses. Strikingly, the replicative protein sequences of the representative virus, barley yellow dwarf virus (BYDV), are closely related to those of carmo- and tobusviruses exploiting A5 strategy (38).

**B7.** This strategy was found in two related groups of plant viruses, which do not seem to show close evolutionary links with other virus groups (39). Importantly, however, the putative proteases mediating polyprotein processing in these viruses are related to the proteases of picorna-, como-, nepo-, and potyviruses (40–42).

Summing up, it is obvious that positive-strand RNA virus genome strategies are not necessarily monophyletic characters. On the other hand, related virus groups tend to have identical or similar strategies. Interdependencies apparently exist between some of the expression and replication mechanisms, the most striking of these being the positive correlation between VPg utilization and polyprotein processing. This “rule”, however, seems to be violated in a subdivision of luteoviruses (B5 strategy).

There are interesting regularities in the host ranges of viruses with different genome strategies. Specifically, all known positive-strand RNA viruses infecting animals produce at least part of their proteins by virus protease-mediated processing of polyprotein precursor(s); on the other hand, genome segmentation is generally not typical of them (nodaviruses being an exception). The opposite trend is apparent in plant viruses, where a constantly growing number of groups is being found to solve their expression problems via subgenomic mRNA genera-

tion and genome segmentation (in the terms of the present classification scheme, this means that strategies A5 and A6 are predominant).

### **Potentially Possible Genome Strategies (Empty Partitions of the Classification Table)**

An exciting possibility inherent in a combinatorial classification like that developed here is the prediction of new combinations that are more or less likely to be realized. For genome strategies, this means that the probability of discovering a yet to be found strategy in a newly studied virus group (i.e., of filling an empty partition in the classification table) can be assessed. It would be of major interest to try to realize what (if any) of the not yet found strategies are "prohibited" by some fundamental principles, and what are likely to be eventually identified. It is to be stated explicitly that no such fundamental "prohibitions" are currently known. On the other hand, we believe that the chances for realization of different strategies are far from equal. First, there are three pairs of actually observed pairs of strategies with and without genome segmentation: A5-A6, A7-A8, and B3-B4. Thus it seems most likely that complementary strategies exploiting segmentation should also exist for strategies A3, B5, and B7, i.e., strategies A4, B6, and B8 are likely to be found in newly characterized virus groups. The situation seems to be somewhat different, with the A2 strategy corresponding to putative RNA bacteriophages with multipartite genomes. The same logic that was forwarded above with respect to the possibility of combination of polycistronic RNA with other expression mechanisms can be applicable here. Specifically, one can argue that, given the ability to utilize multiple ribosome entry sites within a single piece of RNA, genome segmentation would be of no selective advantage. On the other hand, the  $\phi 6$  bacteriophage genome consists of three segments of double-strand RNA, the transcript of each segment being itself a polycistronic RNA (43).

B1 and B2 strategies would correspond to VPg-containing RNA bacteriophages. The search for such phages is a task of major interest, particularly taking into account the existence of the protein-priming mechanism in some DNA bacteriophages (44). On the other hand, the fact that only one group of closely related single-strand RNA bacteriophages has been identified so far, opposing the mirth of DNA bacteriophages, hints at some not yet understood restrictions that might be imposed on RNA replication in bacteria. Exploration of the nature of these possible limitations may be quite intriguing.

Finally, it is of interest to briefly discuss the possibility of finding viruses exploiting the A9 and B9 strategies. Such viruses would express their proteins by the simple principle of "one segment-one protein," without any additional mechanisms. In fact, the putative A9 strategy closely resembles that exploited by tricornaviruses (A6). The only tricornavirus subgenomic mRNA, the one encoding the capsid protein, is incorporated into virions (RNA 4) and is required for infectivity when the infection is initiated with purified virus RNA (45). The only deviation from the putative A9 strategy is that RNA 4 also constitutes the

3' part of RNA 3 and is synthesized as a subgenomic RNA by internally initiated transcription (46). Evidence has been presented that RNA 3 could arise by the fusion of two RNA segments (47). If so, a hypothetical ancestral tricornavirus (or, more precisely, "four"cornavirus) could exploit the A9 strategy. It will be of interest to learn whether this strategy is utilized by some yet unexplored extant virus, and if not, what might be the specific selective advantage conferred by subgenomic mRNA formation. As a matter of speculation, it is possible to propose that such an advantage might lie in the possibility of selective enhancement of subgenomic mRNA transcription required for the abundant production of the capsid protein. Similar considerations could, in principle, apply to the B9 strategy, but its realization seems to be somewhat less likely because of the correlation between VPg-mediated replication and polyprotein processing discussed above.

## Conclusions

The type of classification developed here is obviously not suitable for the purposes of taxonomy. Rather, its potential value lies in the possibility of predicting new combinations of expression and replication mechanisms, and in the explicit formulation of the problems pertaining to the evolution of genome strategies.

## Acknowledgments

I am most grateful to Professor V.I. Agol, who was first to introduce me to the world of RNA viruses. Critical reading of the manuscript by Drs. V.V. Dolja, A.E. Gorbalenya, and T.G. Senkevich is gratefully acknowledged.

## References

1. Matthews R.E.F., *Intervirology* 17, 1–199, 1982.
2. Brown F., *Intervirology* 30, 181–186, 1989.
3. Strauss E.G. and Strauss J.H., *Curr Top Microbiol Immunol* 105, 1–98, 1983.
4. Joshi S. and Haenni A.-L., *FEBS Lett* 177, 163–174, 1984.
5. Zimmer D. in Holland J.J., Domingo E. and Ahlquist P. (eds.) *RNA Genetics*. CRC Press, Boca Raton, FL, 1988, pp. 211–240.
6. Gorbalenya A.E. and Koonin E.V., *Nucleic Acids Res* 17, 8413–8440, 1989.
7. Goldbach R., *Microbiol Sci* 4, 197–202, 1987.
8. Goldbach R. and Wellink J., *Intervirology* 29, 260–267, 1988.
9. Kozak M., *Microbiol Rev* 47, 1–45, 1983.
10. Parker J., *Microbiol Rev* 53, 273–298, 1989.
11. Strauss E.G., Levinson R., Rice C.M., Dalrymple J. and Strauss J.E., *Virology* 164, 265–274, 1988.
12. Kuhn R.J. and Wimmer E. in Rowlands D.J., Mayo M.A. and Mahy B.W.J. (eds.) *The Molecular Biology of Positive Strand RNA Viruses*. Academic Press, London, 1987, pp. 17–51.



13. Tobin G.J., Young D.C. and Flanagan J.B., *Cell* 59, 511–519, 1989.
14. Miller W.A., Bujarski J.J., Dreher T.W. and Hall T.C., *J Mol Biol* 187, 537–546, 1986.
15. Morozov S.Yu., Dolja V.V. and Atabekov J.G., *J Mol Evol* 29, 52–62, 1989.
16. Koonin E.V., Gorbalenya A.E., Chumakov K.M., Donchenko A.P. and Blinov V.M., *Molek Genetika* 7, 27–39, 1987.
17. Koonin E.V., Chumakov K.M., Yushmanov S.Yu. and Gorbalenya A.E., *Molek Genetika* 3, 16–19, 1988.
18. Inokuchi Y., Jacobson A.B., Hirose T., Inayama S. and Hirashima A., *Nucleic Acids Res* 16, 6205–6221, 1988.
19. Collet M.S., Anderson D.K. and Retzel E., *J Gen Virol* 69, 2637–2643, 1988.
20. Gorbalenya A.E., Koonin E.V., Donchenko A.P. and Blinov V.M., *Nucleic Acids Res* 17, 4713–4730, 1989.
21. Gorbalenya A.E., Donchenko A.P., Koonin E.V. and Blinov V.M., *Nucleic Acids Res* 17, 4713–4730, 1989.
22. Carrington J.C., Heaton L.A., Zuidema D., Hilman B.I. and Morris T.J., *Virology* 170, 219–226, 1989.
23. Rochon D.M. and Tremaine J., *Virology* 169, 251–259, 1989.
24. Rupasov V.V., Morozov S.Yu., Kanyuka K.V. and Zavriev S.K., *J Gen Virol* 70, 1861–1869, 1989.
25. Mackenzie D.J., Tremaine J.H. and Stace-Smith R., *J Gen Virol* 70, 1053–1063, 1989.
26. German S., Candresse T., Lanneau M., Huet J.C., Pernollet J.C. and Dunez J., *Virology*, in press, 1990.
27. Xiong Z. and Lommel S.A., *Virology* 171, 543–554, 1989.
28. Hamilton W.D.O., Boccara M., Robinson D.J. and Baulcombe D.C., *J Gen Virol* 68, 2563–2575, 1987.
29. Krausslich H.G. and Wimmer E., *Ann Rev Biochem* 57, 701–754, 1988.
30. Baker S.C., Shieh C.-K., Soe L., Chang M.-F., Vannier D.M. and Lai M.M.C., *J Virol* 63, 3693–3699, 1989.
31. Gorbalenya A.E., Koonin E.V., Donchenko A.P. and Blinov V.M., *Nucleic Acids Res* 17, 4456–4469, 1989.
32. Dominguez G., Wang C.-Y. and Frey T.K., *Virology* 177, 225–238, 1990.
33. Rozanov M.N., Morozov S.Yu. and Skryabin K.G., *Virus Genes* 3, 373–379, 1990.
34. Morch M.-D., Drugeon G., Szafranski P. and Haenni A.L., *J Virol* 63, 5153–5158.
35. Friesen P.D. and Rueckert R.R., *J Virol* 37, 876–886, 1981.
36. Domier L.L., Shaw J.G. and Rhoads R.E., *Virology* 158, 20–27, 1987.
37. Gorbalenya A.E., Blinov V.M., Donchenko A.P. and Koonin E.V., *J Mol Evol* 28, 256–268, 1989.
38. Miller W.A., Waterhouse P.M. and Gerlach W.L., *Nucleic Acids Res* 16, 6097–6111, 1988.
39. Veidt I., Lot H., Leiser M., Scheidecker D., Guilley H., Richards K. and Jonard G., *Nucleic Acids Res* 16, 9917–9932, 1988.
40. Gorbalenya A.E., Koonin E.V., Blinov V.M. and Donchenko A.P., *FEBS Lett* 236, 287–290, 1988.
41. Gorbalenya A.E., Donchenko A.P., Blinov V.M. and Koonin E.V., *FEBS Lett* 243, 103–114, 1989.
42. Bazan J.F. and Fletterick R.J., *FEBS Lett* 249, 5–7, 1989.
43. Mindich L., Nemhauser I., Gottlieb P., Romantschuk M., Carton J., Frucht S., Strassman J., Bamford D.H. and Kalkkinen N., *J Virol* 62, 1180–1185, 1988.
44. Stillman B.W., *Cell* 35, 7–9, 1983.
45. Jaspars E.M.J. in Davies J.W. (ed.) *Molecular Plant Virology*, Vol. 1, CRC Press, Boca Raton, FL, 1985, pp. 155–221.
46. Miller W.A., Dreher T.W. and Hall T.C., *Nature* 313, 68–70, 1985.
47. Ahlquist P., Lukow V. and Kaesberg P.J., *Mol Biol* 153, 23–38, 1981.