J Mol Evol (1987) 25:277–281

Journal of
Molecular Evolution

The Evolution of Multicompartmental Genomes in Viruses

Sean Nee

Department of Biology, University of Sussex, Falmer, Brighton, Sussex BN1 9QG, UK

Summary. The genetic information of many viruses is divided between separately encapsidated nucleic acid molecules. A simple evolutionary model is constructed to explain this phenomenon. All multicompartmental viruses infect plants, and most are RNA viruses. The former fact may be due to the high transmission multiplicities enjoyed by plant viruses. The latter may be due to the low replication fidelity of RNA, although another explanation is also offered. The logic of the analysis is contrasted with that of previous explanations. In particular, this paper proceeds from a "selfish DNA" viewpoint. It is not necessary to suppose that the division of the genome fills any adaptive function for the virus. The theory makes testable predictions about the parameters of multicompartmental viruses.

Key words: RNA viruses — Divided genomes — Copying fidelity — Plant viruses — Defective interfering viruses

Introduction

Many viruses require the genetic information of several nucleic acid molecules for successful replication and/or encapsidation (see Bruening 1977; Van Vloten-Doting and Jaspars 1977; Matthews 1979). In "multicompartment" viruses these molecules are separately encapsidated. Tobacco rattle virus, for example, is a two-component RNA virus. The virions are different sizes; the RNA of the long particle contains the replicase gene while the coat protein gene is carried by the short particle. A cell must be infected by both members of the complementing pair if infectious progeny are to be produced.

A nucleic acid molecule that requires complementation runs the risk that the cell it is in will not be coinfected by its complementing molecule. Hence, an incomplete molecule may suffer a fitness cost (due to its need for complementation) relative to a complete molecule, i.e., one that contains all the information necessary for a complete reproductive cycle.

However, incomplete molecules have a higher replication fidelity than complete molecules. A molecule that codes only one functional gene presents a shorter target to deleterious mutations than a molecule that codes two functional genes.

To fix ideas, consider three kinds of molecules: (1) one codes only a functional coat protein gene; (2) one codes only a functional replicase gene; and (3) one codes both functional coat protein and replicase genes. Call these three kinds CP, R, and CP/R. CP/R has the advantage that it does not require complementation. But, given that they do replicate, CP and R are more likely to produce functional CP and R offspring. CP/R is more likely to suffer deleterious mutation in one or the other of its genes. If the mutations result in complete loss of function, then the mutated offspring of CP/R may be CP or R.

The relationship between the probability of deleterious mutation and complementation probability determines whether multicompartmentalism will evolve.

One implication of the model of this paper is that it is not necessary to suppose that multicompartmentalism fills any function or yields any advantage for the virus. It is simply a stable point in the frequency-dependent evolution of nucleic acid molecules.

Offprint requests to: S. Nee

Previous Explanations

Many explanations of multicompartmentalism have been suggested. Most postulate some advantage or function at the level of the covirus (the complex of mutually complementing defective virions). For example, Van Vloten-Doting and Jaspars (1977) suggested that the phenomenon is a substitute for recombination for RNA viruses [Pressing and Reanney (1984) referenced many occurrences of this idea]. Recombination may be advantageous for a variety of reasons. One especially relevant to RNA viruses is the lowering of "mutation load" (Crow, in press). Other suggested "explanations" can be found in Jaspars (1974) and Bruening (1977). These verbal explanations have been critiqued by Pressing and Reanney (1984).

The only attempt to provide a quantitative evolutionary model of multicompartmentalism is that of Pressing and Reanney (1984). They presented a model in which the replication fidelity of the coviral genome is increased when the genome is divided into separate molecules. This fidelity increase is supposed to provide the selective advantage of multicompartmentalism.

In their model, the fidelity advantage gained is proportionally greater, the larger the genomic replication error rate. This has the virtue of seeming to explain why almost all multicompartmental viruses are RNA viruses. RNA has a per-nucleotide replication error rate of 10^{-3} to 10^{-4}, whereas the rate for DNA is 10^{-9} to 10^{-11}. This difference may be due to the lack of error-correcting mechanisms in enzymes that replicate RNA (Reanney 1982).

Pressing and Reanney (1984) compared the replication fidelity of the *ensemble* of incomplete molecules, the "covirus," to that of a hypothetical complete molecule. But they did not explain why evolution would seek to maximize their measure of the replication fidelity of the covirus ensemble.

Any explanation of multicompartmentalism that invokes some property of the covirus ensemble is a group selectionist explanation. This sort of explanation is relevant if organisms infected by the covirus are not also infected by the complete virus with which natural selection is comparing it. In general, group selection explanations require a very special population structure in order to work.

On the other hand, an analysis that treats the individual replicating molecules as the units of evolution does not necessarily presuppose anything about the natural history of virus transmission. Assumptions about the transmission process may be introduced as desired to see how these assumptions affect the outcome. The following model is such an "individual selection" model.

Model

There are a large number of parameters involved in the evolution of multicompartmentalism. The goal of the model is a graphic representation of viral parameter space with delineated regions within which multicompartmentalism can evolve.

Imagine a virus consisting of two functional genes, a replicase and a coat protein gene. Unconditionally deleterious mutation will give rise to some genomes containing only a functional replicase gene and other genomes containing only a functional coat protein gene. It is assumed that such defective genomes are the components of a mutually complementing two-component system. If a cell is infected by both types of incomplete molecule it will have the complete set of genetic information necessary for the production of potentially infectious virus progeny.

We will define the fitness of a molecule to be the average number of functionally identical copies made from that molecule in an infected cell. Let q be the probability, per nucleotide per replication, that a copying error rendering the gene nonfunctional will *not* occur (so q measures copying fidelity). Let the number of nucleotides in the functional genome of a complete molecule be L, while the number in each of the two complementing incomplete molecules is L/2. The fitnesses of a complete virus and an incomplete molecule, W_C and W_I, are

$$W_C = K_C q^L, \tag{1a}$$

$$W_I = K_I q^{L/2} R. \tag{1b}$$

K is the number of copies of a molecule made within a cell if replication occurs at all. R is the probability that an incomplete molecule will be complemented, i.e., that it will be replicated at all. q^L ($q^{L/2}$) is the replication fidelity of the complete (incomplete) molecules. I assume that R and K are the same for both types of incomplete molecule and that they have the same functional genome length.

Incomplete molecules will certainly increase in frequency if

$$W_I > W_C, \text{ i.e.,} \tag{2a}$$

$$R/q^{L/2} > K_C/K_I. \tag{2b}$$

As the composition of the virus population changes, the Ks will change. This is because the number of copies of a molecule made in a cell depends on the quantities of functional gene product contained therein (see, e.g., Cole and Baltimore 1973). The amounts of functional gene product depend in turn on the characteristics of the infecting viruses; it is assumed in this analysis that the *ratio* K_C/K_I is a constant.

R may also change as evolution proceeds. If it decreases, then although incomplete molecules may

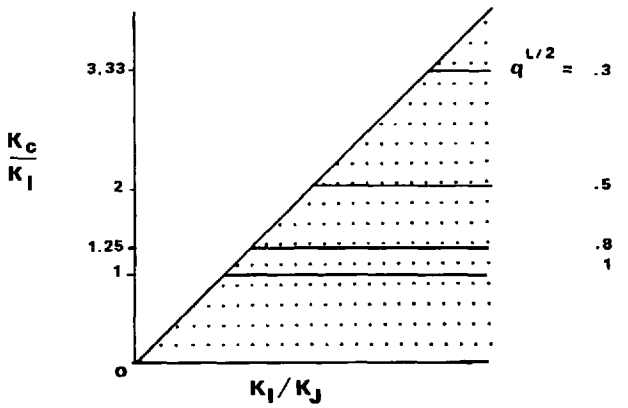

Fig. 1. The dots show the region of $(K_I/K_J, K_C/K_I)$ parameter space for which the fixation condition, $R/q^{L/2} > K_C/K_I$, can possibly be satisfied for some combination of R and $q^{L/2}$. When the condition is satisfied, the complete virus is entirely eliminated from the virus population. Multicompartmentalism cannot possibly evolve in the undotted region of the picture. Each horizontal line corresponds to a particular replication fidelity. The line intersects the K_C/K_I axis at the inverse of this fidelity. For a particular replication fidelity, fixation is only possible in the dotted region below the corresponding line. In this region, multicompartmentalism will evolve if transmission multiplicity, hence R, is sufficiently high.

be able to increase in frequency when rare, they may not be able to replace the complete molecules entirely. That is, they may not be able to go to fixation. Multicompartmentalism has evolved when the incomplete molecules do go to fixation, i.e., when the complete molecules are entirely eliminated from the population. What are the conditions necessary for this to occur?

For fixation, condition (2) must be satisfied when the complete molecules are rare.

It is possible to put an upper limit on the magnitude of $R/q^{L/2}$ according to the following reasoning. When a function-loss mutation occurs in an incomplete molecule, a "junk" molecule is produced that codes for no functional protein at all. A junk molecule may be replicated and encapsidated if it is complemented by *both* types of incomplete molecule, as long as it still has its replicase and encapsidation protein recognition sites.

Recall that the fitness of an incomplete molecule, W_I, is

$$W_I = K_I q^{L/2} R. \tag{3}$$

The fitness of a junk molecule, W_J, is

$$W_J = K_J R^2. \tag{4}$$

This expression assumes that mutations (1) restoring lost function and (2) disrupting recognition sites are so rare that the replication fidelity of junk is approximately one. We also assume an independent probability of being complemented by each type of incomplete molecule.

For the incomplete molecules to be at equilibrium with their derived junk, we must have $W_J < W_I$. (Note that equilibrium does not occur when $W_J = W_I$ because some of the offspring of incomplete molecules are junk molecules.) Hence, a necessary condition for equilibrium is

$$R/q^{L/2} < K_I/K_J. \tag{5}$$

If this condition is not satisfied, then the frequency of junk will rise and complementation probability will fall until the condition *is* satisfied.

From condition (2) we know that for fixation to occur it must be the case that $R/q^{L/2} > K_C/K_I$. From condition (5) we also know that near fixation, $R/q^{L/2} < K_I/K_J$. Therefore, a condition that must be satisfied for fixation to be possible is

$$K_C/K_I < K_I/K_J. \tag{6}$$

This relationship is necessary, but not sufficient, and is satisfied in the dotted region of Fig. 1.

This region of possible fixation may be further partitioned by replication fidelity. Whether or not fixation actually occurs in the dotted region depends on condition (2). Since $R < 1$, fixation of incomplete molecules cannot possibly occur unless $1/q^{L/2} > K_C/K_I$. This requirement is illustrated in Fig. 1, showing that the region in which fixation is possible increases as replication fidelity falls.

Fixation actually *will* occur in the possible region if complementation probability, R, is sufficiently large. R is a function of the proportion of junk in the virus population and the transmission multiplicity, i.e., the average number of viruses that infect a cell. It is shown in the Appendix that for any reasonable R function, increasing transmission multiplicity will stably increase R [up to its maximum, given by (5)]. Because of the way Fig. 1 is constructed, we can say that if a virus's parameters fall in a possible fixation region of Fig. 1, then *the fixation of incomplete molecules will occur if transmission multiplicity is sufficiently high.*

Figure 1 shows that it is not *necessary* that replication fidelity be low for incomplete molecules to go to fixation. In the Discussion we will present an additional explanation of the preponderance of RNA as the genetic material of multicompartmental viruses, one that emphasizes high transmission multiplicity rather than the low replication fidelity of RNA.

In general, transmission multiplicities may, in fact, be low and therefore complementation probability may be well below its maximum. In Fig. 2 we see the effect of transmission multiplicity on fixation. For this figure I chose a particular replication fidelity (0.8) and a random (Poisson) model of complementation. The Appendix presents the mathematics necessary to construct this graph.

Discussion

As Pressing and Reanney (1984) maintained, replication fidelity is certainly an important parameter in the evolution of multicompartmentalism. But fidelity is not the only important parameter, nor even the most important. Complementation probability is very important as well.

Consider the fact that all multicompartmental viruses are *plant* viruses. It may be that features of plants, such as the absence of an immune system, allow viruses to reach much higher transmission multiplicities than can be attained by animal or bacterial viruses. Of course, plants do have systems that may "localize" the infection, confining the virus to a small area around the point of entry (Sela 1981). Such abilities, however, are commonly of an "all or nothing" sort under the control of a single gene. In any case, Matthews (1970) pointed out that the virus concentration attained in successfully infected plants is quite high enough for complementation not to be a problem upon transmission.

Given a very high complementation probability for plant viruses, the predominance of RNA as the genetic material of multicompartment viruses may simply be due to the following fact: most plant viruses are RNA viruses (Primrose and Dunnock 1980).

It is hard to evaluate the relative importance of (1) the low replication fidelity of RNA and (2) the RNA–plant virus connection in explaining the preponderance of RNA as the genetic material of multicompartmental viruses. The low replication fidelity of RNA viruses is the most important explanatory fact if $K_C/K_I > 1$, as we can see from Fig. 1. There is no direct information on the values of these parameters, but there is some suggestive information to be gleaned from the phenomenon of defective interfering viruses.

One plausible scenario for the launch of a virus population on the evolutionary trajectory leading to multicompartmentalism is an increase in the probability of complementation. Under laboratory culture conditions of high-multiplicity serial passage, "defective interfering viruses" have been observed to increase in frequency in all populations of animal virus studied so far. Defective interfering (DI) viruses are typically deletion mutants that interfere with the infection process because they do not produce one or more necessary functional proteins (Huang 1973; Huang and Baltimore 1977).

Huang (1973) presented evidence that these DI viruses increase in frequency because, in the language of this paper, the DI nucleic acids are superior at the intracellular level, i.e., $K_C/K_I < 1$.

It is unknown why this is the case. It may be that they have amplified their replicase recognition sites,

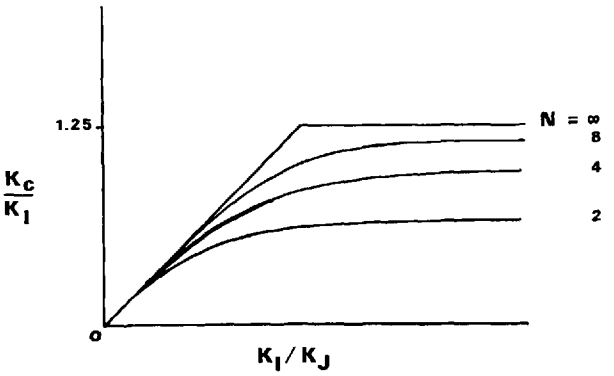

Fig. 2. It is assumed that replication fidelity is 0.8 and that complementation probability, R, is given by $R = 1 - \exp[-N(1 - p)/2]$, where N is the transmission multiplicity and p is the proportion of junk in the population. For a particular transmission multiplicity, fixation will occur in the dotted region below the corresponding curve.

are more attractive to the replicase, or are better at encapsidation. Being shorter may in itself increase replication rate (Speigelman et al. 1975). There are limits, however, to how short a molecule may become, set by the requirements of encapsidation. (In the limit, a molecule achieves an infinite replication rate at a length of zero, but is inefficiently encapsidated.)

Whatever features provide higher K, it is clear that there is a trade-off between having these features and having a complete genome, otherwise the complete molecules would already possess these features. So the DI virus phenomenon weakly suggests that $K_C/K_I < 1$.

One might suppose that the components of a multicompartment virus have incorporated all possible features that provide an edge in the intracellular struggle for survival. This being the case, their derived mutants would be inferior at the intracellular level. So perhaps $K_I/K_J > 1$. I am unaware of any recorded observations of DI viruses in multicompartment systems. This nonobservation is consistent with $K_I/K_J > 1$.

Many plant viruses are not multicompartmental [e.g., the Tymovirus and Tombusvirus groups (Matthews 1979)]. Further natural history information and laboratory information about such groups would be of interest. Perhaps they are not multicompartmental because the probability of complementation is too low because of the natural history of their transmission, or perhaps their other parameters are inappropriate.

Acknowledgments. I thank J. Maynard Smith, A Pomiankowski, J. Pressing, and B. Charlesworth for their valuable input. I thank two anonymous referees for their constructive suggestions and their insistence on comprehensibility.

Appendix

In this Appendix, we will study the complementation probability when complete molecules are rare and incomplete molecules are at equilibrium with their junk. This is called the "equilibrium complementation probability."

Let p be the frequency of junk. The recursion equation is

$$p' = \frac{pK_JR^2 + (1 - p)K_I(1 - q^{L/2})R}{(1 - p)K_IR + pK_JR^2}. \quad \text{(A1)}$$

After some manipulation, this becomes

$$p' = 1 - q^{L/2}i/(i + uR), \quad \text{(A2)}$$

where $i = K_I/K_J$ and $u = p/(1 - p)$. At equilibrium $p' = p$, so

$$\hat{p} = i(1 - q^{L/2})/(i - R). \quad \text{(A3)}$$

This condition defines a line of potential equilibria in (p, R) space.

Notice that in order to satisfy $0 < \hat{p} < 1$, we must have $R < q^{L/2}K_I/K_J$. This necessary condition for equilibrium was derived more directly in expression (5) in the text.

Complementation probability, R, is a function of p and transmission multiplicity, N. For a given N, R is a function of p. If R is a "reasonable" function, it is a declining function of p.

Equilibrium R, $\hat{R}$, is determined by the intersection of R and the line of equilibria (A3). This is illustrated in Fig. A1. This figure also shows that for reasonable R functions, $\hat{R}$ is higher with higher transmission multiplicities.

Given $q^{L/2}$ and N, $\hat{R}$ is a function of K_I/K_J. The curves in Fig. 2 are of $R/q^{L/2}$ as a function of K_I/K_J. They were obtained as follows.

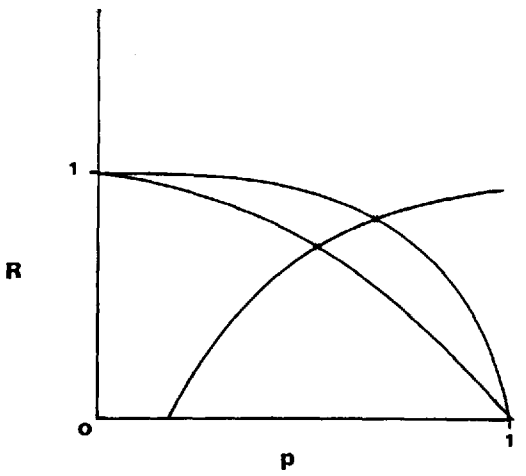

Fig. A1. The increasing line is the line of possible equilibria from equation (A3). The decreasing lines are graphs of R as a function of p, given a particular transmission multiplicity. The higher R has a higher transmission multiplicity. The intersection points are stable equilibria.

Rearranging (A3) we get

$$i = \hat{R}\hat{p}/(q^{L/2} - 1 + \hat{p}). \quad \text{(A4)}$$

If virus particles infect cells randomly,

$$\hat{R} = 1 - \exp[-N(1 - \hat{p})/2]. \quad \text{(A5)}$$

Then

$$\hat{p} = 1 + 2\ln(1 - \hat{R})/N. \quad \text{(A6)}$$

Substitute this value for $\hat{p}$ into (A4) to get

$$\begin{aligned} i &= K_I/K_J \\ &= \hat{R}[1 + 2\ln(1 - \hat{R})/N]/[q^{L/2} + 2\ln(1 - \hat{R})/N]. \end{aligned} \quad \text{(A7)}$$

For $0 < \hat{R} < q^{L/2}K_I/K_J$, K_I/K_J is a strictly monotonic function of $\hat{R}$. Therefore from the set of points $(\hat{R}, K_I/K_J)$ we obtain the set $(K_I/K_J, \hat{R}/q^{L/2})$, which belongs to $\hat{R}/q^{L/2}$ as a function of K_I/K_J.

References

Bruening G (1977) Plant covirus systems: two-component systems. In: Fraenkel-Conrat H, Wagner RR (eds) Comprehensive virology, vol 11. Plenum, New York, pp 55–142

Cole CN, Baltimore D (1973) Defective interfering particles of poliovirus. III. Interference and enrichment. J Mol Biol 76: 345–361

Crow JF (in press) The importance of recombination. In: Michod R, Levin B (eds) The evolution of sex: an examination of current ideas. Sinauer, Sunderland MA

Huang AS (1973) Defective interfering viruses. Annu Rev Microbiol 27:101–117

Huang AS, Baltimore D (1977) Defective interfering animal viruses. In: Fraenkel-Conrat H, Wagner RR (eds) Comprehensive virology, vol 10. Plenum, New York, pp 73–116

Jaspars EMJ (1974) Plant viruses with a multipartite genome. Adv Virus Res 19:37–149

Kassanis B (1968) Satellitism and related phenomena in plant and animal viruses. Adv Virus Res 13:147–180

Matthews REF (1970) Plant virology. Academic Press, New York, p 253

Matthews REF (1979) Classification and nomenclature of viruses. Intervirology 12:150–287

Pressing J, Reanney DC (1984) Divided genomes and intrinsic noise. J Mol Evol 20:135–146

Primrose SB, Dunnock NJ (1980) Introduction to modern virology. Blackwell, Oxford, p 44

Reanney DC (1982) The evolution of RNA viruses. Annu Rev Microbiol 36:47–73

Sela I (1981) Plant–virus interactions related to resistance and localization of viral infections. Adv Virus Res 26:201–237

Spiegelman S, Mills DR, Kramer FR (1975) The extracellular evolution of structure in replicating RNA molecules. In: Miller IR (ed) Stability and origin of biological information. Halsted, New York

Van Vloten-Doting L, Jaspars EMJ (1977) Plant covirus systems: three component systems. In: Fraenkel-Conrat H, Wagner RR (eds) Comprehensive virology, vol 11. Plenum, New York

Received August 11, 1986/Revised February 27, 1987