PROCEEDINGS — OF — **THE ROYAL SOCIETY** **B** | **BIOLOGICAL SCIENCES**

Evolutionary dynamics of genome segmentation in multipartite viruses

Jaime Iranzo and Susanna C. Manrubia

Supplementary data	"Data Supplement" http://rspb.royalsocietypublishing.org/content/suppl/2012/06/27/rspb.2012.1086.DC1.html
References	**This article cites 37 articles, 13 of which can be accessed free** http://rspb.royalsocietypublishing.org/content/early/2012/06/27/rspb.2012.1086.full.html#ref-list-1
P<P	Published online 4 July 2012 in advance of the print journal.
Subject collections	Articles on similar topics can be found in the following collections evolution (1230 articles) theoretical biology (30 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

Evolutionary dynamics of genome segmentation in multipartite viruses

Jaime Iranzo and Susanna C. Manrubia*

Centro de Astrobiología, INTA-CSIC, Ctra. de Ajalvir km. 4, 28850 Torrejón de Ardoz, Madrid, Spain

Multipartite viruses are formed by a variable number of genomic fragments packed in independent viral capsids. This fact poses stringent conditions on their transmission mode, demanding, in particular, a high multiplicity of infection (MOI) for successful propagation. The actual advantages of the multipartite viral strategy are as yet unclear. The origin of multipartite viruses represents an evolutionary puzzle. While classical theories suggested that a faster replication rate or higher replication fidelity would favour shorter segments, recent experimental results seem to point to an increased stability of virions with incomplete genomes as a factor able to compensate for the disadvantage of mandatory complementation. Using as main parameters differential stability as a function of genome length and MOI, we calculate the conditions under which a set of complementary segments of a viral genome would outcompete the non-segmented variant. Further, we examine the likeliness that multipartite viral forms could be the evolutionary outcome of the competition among the defective genomes of different lengths that spontaneously arise under replication of a complete, wild-type genome. We conclude that only multipartite viruses with a small number of segments could be produced in our scenario, and discuss alternative hypotheses for the origin of multipartite viruses with more than four segments.

Keywords: segmented viral genomes; multiplicity of infection; complementation; mathematical model

1. INTRODUCTION

The origin and evolutionary history of viral genomes is a classical problem that has inspired a long series of questions and hypotheses in evolutionary biology [1–3]. One of those questions is the adaptive meaning of genome segmentation, because it appears to be a common trait in a very broad variety of viruses [4–6]. The case of multipartite viruses is particularly striking, because their genome segments achieve complete independence at the apparent cost of reducing its infectivity [7–9].

Multipartite viruses have their genomes fragmented into two or more (up to eight) segments, each packed into a separate capsid and containing one or more genes that are essential for the virus to complete an infection cycle. These viruses require complementation, because each genomic segment must be completed (i.e. complemented) with the rest of segments in order to produce viral offspring. The complementation requirements for multipartite viruses have a strong impact on the way they are transmitted: many viral particles have to enter each cell in order to assure that at least one representative of each segment will be present. The multiplicity of infection (MOI) is thus a key quantity in the biology of multipartite viruses.

A noticeable fact about multipartite viruses is their asymmetry in host distribution: while they are common among plant viruses, no multipartite virus infecting animals has been described [10]. It has often been claimed that a larger characteristic MOI in plant infections may be behind this phenomenon [11], but a quantitative analysis

of this claim has not been carried out up to now. From an evolutionary perspective, this asymmetry could be understood by taking into account a trade-off between opposite selective pressures: while the complementation requirement acts as a limiting factor—viral extinction supervenes if transmission bottlenecks occur—there should be evolutionary forces that promote the fragmentation of viral genomes. Actually, the first step towards fragmentation might be the generation of incomplete genomes. The latter are known to arise spontaneously under replication of wild-type (wt) viruses [12], which often produce the so-called defective interfering particles (DIPs)—that is, incomplete genomes able to infect cells but unable to complete the infection cycle in the absence of the wt [13,14]. The effect of DIPs in the dynamics and evolution of viruses has been studied by means of mathematical models [15,16], and particular attention has been paid to the mechanisms allowing for the coexistence of wt and defective forms [4,17,18]. Nonetheless, the advantage of fragmentation and especially the individual encapsidation of the fragments still remain open questions.

Faster replication of shorter genomes and higher replication fidelity have been classically presented as factors favouring genome segmentation [11,19], though there is no conclusive empirical evidence of their evolutionary advantage in viruses. Recent experimental work, on the other hand, has compared the performance of the wt, complete form of foot-and-mouth disease virus, with a fragmented (bipartite) counterpart obtained by evolution in cell culture [20]. Competition experiments have shown that the latter, shorter genomes, may possess a larger average lifetime between infective events [6]. These results point at the stability of viral particles as a relevant feature that could counterbalance the disadvantage of high MOIs required to produce infection.
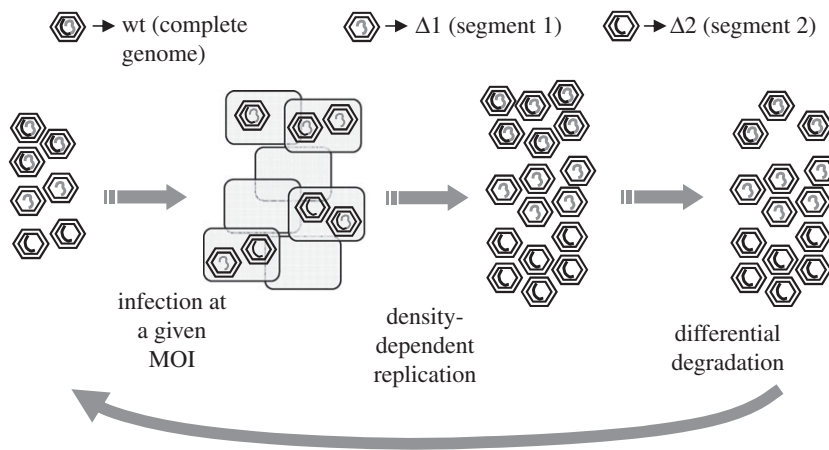
* Author for correspondence (scmanrubia@cab.inta-csic.es).

Figure 1. Schematic of the infective process. The MOI is the (average) number of viral particles infecting a cell. Incomplete genomes ($\Delta1$ and $\Delta2$) have an average lifetime that is larger than that of the wt between infective events, as shown graphically by the decrease in number of the wt relative to genomes with deletions. The process is iterated until the equilibrium state is attained.

While the possibility of obtaining complementary, segmented variants of an originally non-segmented virus had been confirmed through isolation of variants with genomes separated into two molecules [21] and by means of genetic engineering techniques [22,23], the experimental evolution of a bipartite virus from a complete, non-segmented wt provides an insight into how multipartite viruses could have originated in nature. In this work, we explore the hypothesis that multipartite viruses are the evolutionary outcome of the competition among genomic segments of different lengths. These segments would be naturally produced through deletions in the replication process of the original, wt virus [24]. Provided that shorter genomes enjoy a certain evolutionary advantage, a set of segments may be able to outcompete the wt virus if the MOI is high enough to guarantee complementation among the segments. We will focus on reduced degradation as the selective pressure favouring segmentation, although alternatives will also be considered, and their formal equivalence investigated. We discuss the likeliness that multipartite viruses with a large number of fragments could have originated in that scenario.

2. MATERIAL AND METHODS

(a) *Formal scenario*
A schematic of the basic mechanisms included in the model is depicted in figure 1. Let us consider a mixed population that contains wt viruses (those with a complete genome, denoted as wt) as well as two smaller components whose genomes are complementary segments (denoted as $\Delta1$ and $\Delta2$). Both components $\Delta1$ and $\Delta2$ constitute together a bipartite variant of the wt virus. Each component requires complementation for replication, either with the other component or with the wt. At each generation, viruses in the population infect a set of cells at a given MOI m. Although, in general, m is a quantity that depends on the viral load, we will study only the case of constant m for the sake of simplicity. Replication takes place inside the cells depending on complementation requirements. Thus, in the case of frequency-dependent fitness [25], the amount of viruses of each class produced by a single cell depends on the initial composition of the (small) infecting population. The sum of all viruses produced

by all cells constitutes the primary offspring. It is exposed then to differential degradation, which preferentially affects the wt class. Survival of the wt relative to the segmented classes is given by a parameter $\sigma < 1$. The viral population that results after degradation is considered to be the infecting population for the next generation. In such a way, the composition of the population can be traced for several generations in an iterative way.

(b) *Evolution equation with differential degradation*
The composition of the population can be expressed by a vector

$$\mathbf{p} = (p_{\Delta1}, p_{\Delta2}, p_{wt})^T \tag{2.1}$$

where $p_{\Delta1}$, $p_{\Delta2}$ and p_{wt} are the fractions of classes $\Delta1$, $\Delta2$ and wt in the population, such that $p_{\Delta1} + p_{\Delta2} + p_{wt} = 1$. At generation n, the composition of the population will be denoted as $\mathbf{p}_n$.

Let us begin by considering a single cell that has been infected by a, b and c viral particles of classes $\Delta1$, $\Delta2$ and wt, respectively. The MOI in this case would be $m = a + b + c$. Vector $(a, b, c)^T$ will be referred to as the *infection configuration*. The viral offspring produced by the infected cells can be expressed as a product of a fitness matrix $\mathbf{M}_{a,b,c}$ and the infection configuration. The fitness matrix allows the introduction of frequency-dependent fitness, provided that production of each viral class is affected by the abundances of other classes in a linear way.

The complementation requirement can be implemented as follows. For a given genome to reproduce, there are two limiting factors: first, the availability of the own genome; second, the availability of essential proteins (that depends itself on the amount of genomes coding for them). As a result, a simple way to consider complementation consists of assuming that the number of genomes of a given class produced inside a cell is equal to the minimum between the number of genomes of that class that infected the cell and the number of the corresponding complementary genomes. For instance, segment $\Delta1$ requires $\Delta2$ or wt for complementation; so the number of viral particles of class $\Delta1$ produced will be $\min\{a, b + c\}$. On the other hand, wt genomes do not need complementation, and their final number will depend only on their initial abundance c. Without loss of

generality, we will fix the replication rate for the wt class equal to one. Taking all this into account, the offspring produced by a single cell can be written as

$$\mathsf{M}_{a,b,c}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \min\{a, b+c\} \\ \min\{b, a+c\} \\ c \end{pmatrix} \equiv \begin{pmatrix} f_{\Delta 1|a,b,c} \\ f_{\Delta 2|a,b,c} \\ \sigma^{-1} f_{\mathrm{wt}|a,b,c} \end{pmatrix}. \qquad (2.2)$$

The previous expression defines conditional (frequency-dependent) fitness $f_{i|a,b,c}$, $i \in \{\Delta 1, \Delta 2, \mathrm{wt}\}$.

The next step is to obtain the global offspring produced by the whole set of cells. Let us assume that the number of cells is large enough such that the global offspring can be calculated by averaging the offspring of a single cell (equation (2.2)) over the probability of a given infection configuration. That probability depends on the MOI m as well as on the population composition $\mathbf{p}$. As an example, we consider here the case where the MOI is Poisson distributed,

$$\mathrm{Pr}(a,b,c|\mathbf{p}) = e^{-m} \frac{m^{(a+b+c)}}{a!b!c!} p_{\Delta 1}^{a} p_{\Delta 2}^{b} p_{\mathrm{wt}}^{c}. \qquad (2.3)$$

Note that this joint distribution is equivalent to the product of three independent Poisson distributions with averages mp_i, with $i \in \{\Delta 1, \Delta 2, \mathrm{wt}\}$. The case of an MOI following a multinomial distribution is dealt with in the electronic supplementary material.

Differential degradation is applied by multiplying the global offspring by a degradation matrix D that takes the form of a diagonal matrix with value one in the first two positions (corresponding to $\Delta 1$ and $\Delta 2$) and $\sigma < 1$ in the third position (reduced survival for the wt),

$$\mathsf{D} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sigma \end{pmatrix}. \qquad (2.4)$$

All steps can be written in a single equation that provides the composition of the population in successive generations.

$$\mathbf{p}_{n+1} = Z^{-1} \mathsf{D} \sum_{a,b,c} \mathrm{Pr}(a,b,c|\mathbf{p}_n) \mathsf{M}_{a,b,c} \begin{pmatrix} a \\ b \\ c \end{pmatrix} \qquad (2.5)$$

where Z is a normalization factor (see the electronic supplementary material).

Equation (2.5) can be iterated in order to obtain the evolution of a mixed population of single-particle wt virus and bipartite mutants derived from it, provided that the selective advantage of the bipartite mutants is due to reduced degradation. Note that according to the definition of conditional fitness (equation (2.2)), it can be written in the form of a replicator equation. Indeed, let $\langle f_i \rangle$ be the average value of the conditional fitness for a generic class i,

$$\langle f_i \rangle = \sum_{a,b,c} \mathrm{Pr}(a,b,c|\mathbf{p}) f_{i|a,b,c}. \qquad (2.6)$$

If p_i is the fraction of class i in the population and we denote with a superscript the generation in which a given quantity is measured, the evolution equation is equivalent to the replicator equation

$$p_i^{(n+1)} = \frac{\langle f_i \rangle^{(n)}}{Z^{(n)}} \quad i \in \{\Delta 1, \Delta 2, \mathrm{wt}\}, \qquad (2.7)$$

where

$$Z^{(n)} = \sum_i \langle f_i \rangle^{(n)}. \qquad (2.8)$$

The evolutionary dynamics reaches an equilibrium state for such compositions that are fixed points of the replicator equation. They must fulfil the equilibrium condition

$$p_i^* = \frac{\langle f_i(\mathbf{p}*) \rangle}{Z} \quad \forall i \in \{\Delta 1, \Delta 2, \mathrm{wt}\}. \qquad (2.9)$$

In particular, we are interested in equilibrium points that are attractors of the evolutionary dynamics. A detailed study of these points and their stability can be found in the electronic supplementary material.

(c) Generalization of the evolution equation

The evolution equation in the previous section can be modified to include additional selective pressures. The set of fitness matrices will contain new parameters accounting, for instance, for replication and mutation rates, and in some cases, the number of replication cycles inside the cell has to be explicitly considered. A detailed derivation of generalized equations can be found in the electronic supplementary material.

(i) Different replication rates

Owing to their smaller size, it has been argued that bipartite genomes may replicate faster than wt ones [11,19]. Without loss of generality, let $R > 1$ be the replication rate of the segments $\Delta 1$ and $\Delta 2$ relative to that of the wt. In a discrete-time model, R^{-1} is the average number of genomes of wt class produced after one cycle of intracellular replication. Several replication cycles can take place before the viral offspring is released out of the cells, let G be the number of replication cycles. Under these circumstances, the conditional fitness that allow the evolutionary process to be written as a replicator equation are the following:

$$\left. \begin{array}{l} f_{\Delta 1|a,b,c} = \min\{a, b + cR^{1-G}\}, \\ f_{\Delta 2|a,b,c} = \min\{b, a + cR^{1-G}\} \\ \text{and} \quad f_{\mathrm{wt}|a,b,c} = \sigma c R^{-G}. \end{array} \right\} \qquad (2.10)$$

(ii) Loss of segments through mutation and replication fidelity

Let us consider the possibility that a genomic segment is lost during replication with probability ρ. As a result, the probability that a wt virus replicates its whole genome without losing any segment is $(1-\rho)^2$, where the square means that a genome with two putative segments is considered. On the other hand, the probability that a single-segment virus $\Delta 1$ or $\Delta 2$ replicates without errors is $1 - \rho$. That is the reason why bipartite mutants have, in principle, a higher probability of error-free copy than wt ones. In addition, single-segment viruses can be produced from the wt with probability $\rho(1-\rho)$. In this setting, the evolution of the population follows a replicator equation where conditional fitness can be expressed as

$$\left. \begin{array}{l} f_{\Delta 1|a,b,c} = (1-\rho)^G [\min\{a, b + c(1-\rho)^{1-G}\} \\ \qquad\qquad + c(1 - (1-\rho)^G)], \\ f_{\Delta 2|a,b,c} = (1-\rho)^G [\min\{b, a + c(1-\rho)^{1-G}\} \\ \qquad\qquad + c(1 - (1-\rho)^G)] \\ \text{and} \quad f_{\mathrm{wt}|a,b,c} = \sigma c(1-\rho)^{2G} \end{array} \right\} \qquad (2.11)$$
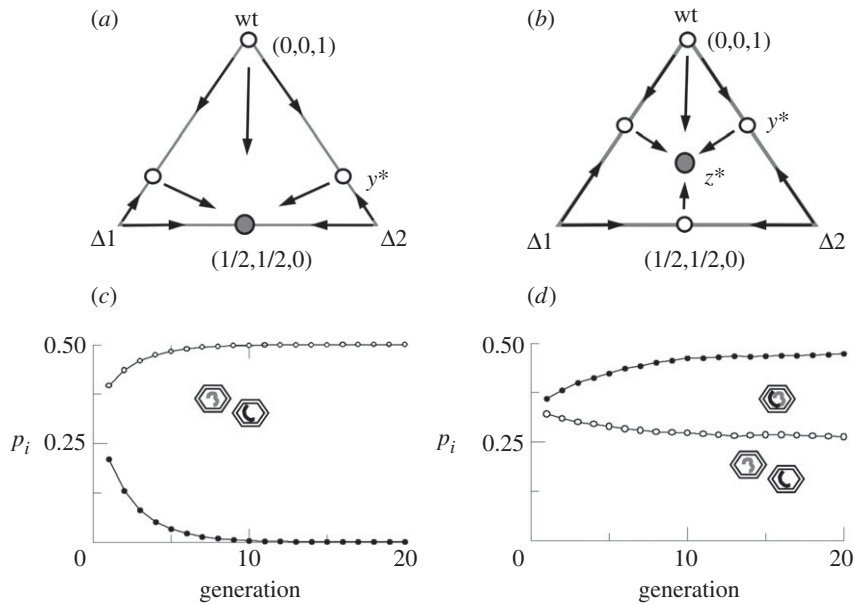
Figure 2. Evolutionary outcomes for the competition between a single particle virus (wt) and its bipartite variant ($\Delta 1$ and $\Delta 2$). (*a,b*) Equilibrium points in the population simplex (open circles, unstable; filled circles, stable); arrows indicate the evolutionary trajectories. (*c,d*) Temporal evolution of population composition for a characteristic realization (filled circles: wt, open circles: $\Delta 1$ and $\Delta 2$, legend for each viral form as in the previous figure). (*a,c*) Fixation of the bipartite variant takes place if survival of the wt is below a critical value $\sigma < \sigma_{\mathrm{crit}}$. (*b,d*) Coexistence of all viral classes occurs if $\sigma > \sigma_{\mathrm{crit}}$. An analytical expression for σ_{crit} is given in the main text.

(iii) *Constant per-cell viral productivity*

The expressions in equation (2.2) implicitly assume that there are no restrictions to the maximal number of viral particles that an infected cell can produce. However, because cellular resources are limited, viral production may be bound. This situation can be tackled by normalizing the total number of viral particles produced. Without loss of generality, we suppose that this value equals 1 and study the dynamics under the conditional fitness

$$\left.\begin{aligned} f_{\Delta 1|a,b,c} &= z_{a,b,c}^{-1}\min\{a, b+c\}, \\ f_{\Delta 2|a,b,c} &= z_{a,b,c}^{-1}\min\{b, a+c\} \\ f_{\mathrm{wt}|a,b,c} &= z_{a,b,c}^{-1}\sigma c, \end{aligned}\right\} \qquad (2.10)$$

and

with $z_{a,b,c} = \min\{a+b, 2a+c, 2b+c\} + c$.

(d) *Multiple segments*

The model can be easily expanded to describe the dynamics of genomes that are susceptible to being partitioned into more than two segments. In a multiple-segment model, viral classes are defined by the genomic segments they conserve. The extreme cases are still the wt virus, which contains the complete genome, and the single-segment classes, which constitute the genuine multipartite version of the virus. In addition, there will also be classes with an intermediate number of segments. Provided that a genome is composed of n putative segments, the total number of different viral classes is $2^n - 1$ (the class containing no segments has been discounted), and the number of classes containing s segments is $\binom{n}{s} = n!/(s!(n-s)!)$.

Complementation implies that replication is limited by the less abundant genomic segment inside the cell. In the simplest scenario, we assume that the selective advantage favouring shorter genome lengths is proportional to the number of segments that a given class contains. In the case of degradative advantage, degradation is assumed to be zero for the single-segment classes and to increase linearly with the number of segments up to the value $1 - \sigma$ for the wt virus. By taking these considerations into account, the extension of the bipartite model to the multipartite case is straightforward, the main difficulties arising from the high dimensionality of the classes space. The case with three segments is developed in detail in the electronic supplementary material. Other relationships between genome length and degradative advantage are possible: the case where the selective advantage is proportional to the volume of the packed genome (emphasizing the role played by the interaction with the capsid) is studied in the electronic supplementary material.

3. RESULTS

(a) *Evolutionary shift from wt to a bipartite form*

Analysis of the evolution equation (2.5) or the equivalent replicator equation (2.7) reveals two possible outcomes for the evolutionary process, depending on the values of parameters σ and m. If degradation of the wt virus is high enough when compared with that of the segments, then the wt becomes extinct and single-segment variants $\Delta 1$ and $\Delta 2$ take over the population (figure 2*a*). Therefore, in this regime bipartite variants of a single-particle virus are able to outcompete the latter and reach fixation in the population. On the other hand, coexistence is the expected outcome if degradation of the wt is low (figure 2*b*). Both regimes are separated by a critical value of the survival parameter σ_{crit}, so that $\sigma < \sigma_{\mathrm{crit}}$ leads to extinction of the wt while $\sigma > \sigma_{\mathrm{crit}}$ allows for coexistence.

An analytic expression for the critical value σ_{crit} can be obtained by means of simple invasibility arguments. Provided that the point $(1/2, 1/2, 0)^T$ (corresponding to a pure equilibrated population of the bipartite form) is a stable equilibrium point in the absence of the wt class, the key point is to study its stability when an infinitesimal
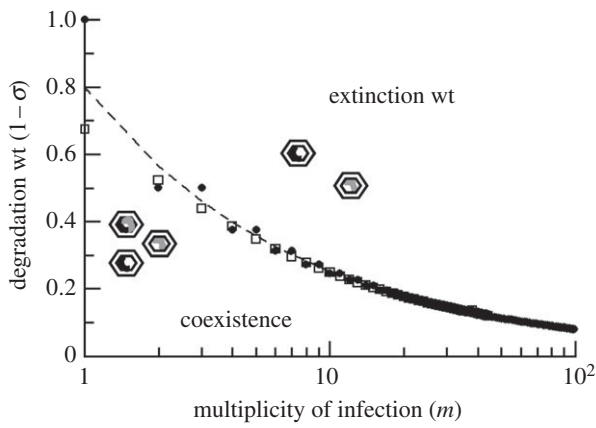
Figure 3. Evolutionary regions depending on selective pressures (MOI and degradation of wt class). Regions corresponding to coexistence and extinction of wt (fixation of bipartite virus) are separated by a series of critical points, whose values vary slightly according to the probability distribution that governs the infection process (solid circles, multinomial; open squares, Poisson, dashed line, asymptotic behaviour).
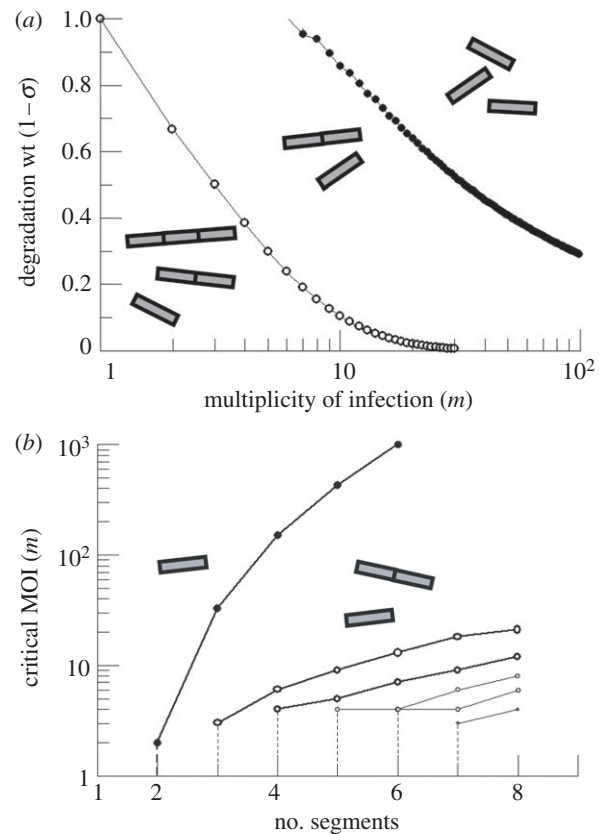

Figure 4. Evolutionary outcomes for genomes with multiple segments. (a) Coexistence and multipartite fixation regions for three segments. Rectangles schematically indicate the number of segments that viral classes contain in a certain region (all classes with such a number of segments will be present). Compare these curves with σ_{crit} for two segments, shown in figure 3. (b) Critical value of the MOI required for fixation of single-segment classes (thick line on the left), for $\sigma = 0.5$. Other lines indicate further transitions: from double-single coexistence to triple-double-single, and so on.

amount of wt is introduced. The equilibrium point becomes unstable at a critical value $\sigma = \sigma_{crit}$, where

$$\sigma_{crit} = \frac{2}{m} \sum_{a,b} \Pr\left(a, b \mid \tfrac{1}{2}, \tfrac{1}{2}, 0\right) \min\{a, b\}. \quad (3.1)$$

Using equation (2.3) and after some algebra (see the electronic supplementary material), the critical value separating coexistence of all types from extinction of the wt can be written in terms of modified Bessel functions of the first kind $I_\alpha(m)$:

$$\sigma_{crit} = 1 - e^{-m} I_0(m) - e^{-m} I_1(m). \quad (3.2)$$

A simple expression is obtained in the limit of large MOI, $m \gg 1$,

$$\sigma_{crit} \sim 1 - \sqrt{\frac{2}{\pi}} m^{-1/2}, \quad (3.3)$$

and this same asymptotic result holds for a multinomial distribution of MOIs.

Figure 3 compares the numerical and asymptotic values of σ_{crit} as a function of m, and reveals that the asymptotic approximation actually recovers well the behaviour of the system also at relatively small values of m. As intuitively expected, increasing the relative degradation of wt virus implies increasing the selective pressure favourable to the bipartite virus, which permits its fixation. Alternatively, an increase of the MOI makes complementation easier, as there are more genomes inside the cell providing complementation. As a consequence, greater MOI also favours fixation of the bipartite virus. Finally, note that there is no parameter region for which the wt virus outcompetes the bipartite one.

The results obtained earlier remain qualitatively unchanged if the selective advantage of the bipartite virus relies on a faster replication or if mutations leading to the loss of segments are considered. In the former case, the conditional fitness defined by equation (2.10) can be used to derive an analogous critical condition such that σ_{crit} is replaced by R^{-G}. In the latter, σ_{crit} is

substituted by $(1 - \rho)^G$. When resources are limited by the cell, the critical condition is derived from a slightly more involved mathematical relationship. Nonetheless, there are no qualitative differences with the situation reported, but only minor quantitative differences (see the electronic supplementary material).

(b) *Viruses with multiple segments*

As in the bipartite case, evolutionary outcomes include coexistence of all possible viral classes and fixation of the single-segment classes. The latter would result in a net evolution from a single-particle virus to a multipartite virus with as many particles as genomic segments. As a novelty, there appears a whole range of intermediate equilibrium states that successively lack the wt, the second longest classes and so on. Figure 4a shows a map of the evolutionary regions for a genome with three possible segments. The intermediate region corresponds to an equilibrium state that contains the three possible two-segment classes as well as the three single-segment classes. This region is limited by two series of critical points that separate it from the total coexistence region below and the multipartite fixation region above. A comparison with the two-segment case (figure 3) reveals that fixation of a multipartite virus with three segments requires a much

higher MOI. This result is expected because, in this case, complementation of a single segment requires the presence of two different complementary segments. For low values of the MOI, it may be impossible for the single segments to get fixation, even with the maximum degradative advantage. This is because of the linear relation between the degradative advantage and the number of segments: the minimum survival for the wt virus, 0 per cent related to that of a single segment, translates into a relative survival of 50 per cent for two-segment classes. That can be enough for the two-segment classes to avoid extinction at a not very high MOI.

A relevant question is how high the MOI must be so that a multipartite virus with a given number of segments can reach fixation. To address that point, let us take a fixed value for the survival parameter and observe the critical MOI values that separate one evolutionary region from another. Results for $\sigma = 0.5$ are shown in figure 4b. The thick black curve on the left-hand side of the figure indicates the minimum MOI required so that the single-segment classes are able to outcompete other classes with longer genomes and, in consequence, a fully multipartite population is established. That critical MOI rapidly increases with the number of segments in the genome, being equal to two for two segments, around 30 for three segments and higher than 100 for four and more segments. If the selection pressure favouring shorter genomes is stronger, let us take for instance $\sigma = 0.1$ for the survival of the wt, then critical MOIs decrease (two for two segments, nine for three segments), but still remain above 100 for five or more segments. The possibility that such a high MOI can be reached in nature is clearly unrealistic (not so many viral particles are expected to enter one cell under biological conditions); therefore, if the evolutionary origin of multipartite viruses is due to genome segmentation and competition among genomes of different lengths, no multipartite virus with more than three or four segments should be expected. This result remains unchanged if the advantage of segmented forms is proportional to the volume they occupy, as shown in the electronic supplementary material.

4. DISCUSSION

We have presented a model of how viral genomes may become segmented and give rise to a multipartite virus. Provided that mutants with shorter genomes can be produced through deletion events and that these mutants are able to replicate if they receive complementation, evolution of a multipartite variant of the original virus may be the result of the competition among genomes with different lengths. Two opposite selective pressures determine whether the multipartite virus will be able to reach fixation. On the one hand, complementation among segmented genomes requires co-infection of cells by at least one segment of each class. That is possible only if the MOI is high enough. On the other hand, shorter genomes benefit from a reduced degradation rate, which favours division of the genome into smaller segments [6]. Other hypothetical benefits of segmentation, such as faster replication or higher fidelity in the replication process have also been considered, with no qualitative changes in the overall results. The main evolutionary regions (coexistence of wt and fragmented forms and extinction of the

wt) are essentially unchanged when the production of viral particles is limited by the amount of cellular resources. This leads us to conclude that it is how the complementation rules are implemented that eventually determines the equilibrium state, and not the absolute number of viral particles produced. At the biochemical level, the expression used to evaluate the fitness of the different infection configurations implicitly assumes that gene products affected by segmentation are partially shared, though the genome coding them can use them preferentially. That is, gene products act in *trans* and partly in *cis*. Other models for complementation [26] have analysed the case of gene products acting only in *trans*. Compared with our scenario, this latter prescription confers a larger advantage to segmented forms. Hence, segmented genomes would be fixed at lower values of MOI, other parameters being equal.

The case with two genomic segments giving rise to a bipartite virus has been solved in a comprehensive, analytical way. It shows that an evolutionary shift from a single-particle virus to a bipartite one is the expected outcome when the MOI during evolution exceeds a critical value. This result explains, from a theoretical point of view, the experimental observations in [20], where a bipartite virus was obtained after culture of foot-and-mouth disease virus (a single-particle virus) at a high MOI.

Our results for the bipartite case can be compared with those obtained in previous theoretical scenarios. In a complementation model characterized by hyperbolic growth [4], both coexistence and fixation of the bipartite form were found, together with a third regime where the wt virus cannot be invaded by segmented mutants. The assumption of a hyperbolic viral growth is essential for this third regime to exist. However, hyperbolic viral growth requires a tight spatio-temporal coupling between synthesis of replication proteins and genome replication, a situation that is not representative for many viruses [27,28]. A related work [19] considered differences in replicative ability and replication fidelity as the selective pressures driving genome segmentation. Interestingly, the model also considered an additional class of parasitic-like, defective mutants that need complementation by single-segment genomes to be replicated and provide no complementation to the rest of the population. For high enough MOI, it was found that defective mutants were able to invade a population of bipartite virus, driving it to extinction. This phenomenon is conceptually similar to that of *lethal defection* [29,30] and has not been considered in our model for simplicity.

Finally, we have extended our study to the case of multiple (more than two) segments. The main result is that a multipartite virus with a small number of segments can outcompete the single-particle one and get fixed in the population at realistic values of the MOI. According to experimental assays, the MOI in plant infections oscillates between 2 and 13 [9,31], which would allow selection of multipartite viruses with two and three segments. For a greater number of segments, total segmentation at realistic MOIs should not be expected in the framework of our model.

(a) *Multipartite viruses are found only in plants*
It is known from experimental assays that the need for complementation reduces the infectivity of multipartite

viruses to efficiencies below 10 per cent [9]. One can calculate the probability that a multipartite virus with n segments achieves complementation when infecting cells at MOI m. If we require that this probability reaches a level of 0.1 (one out of 10 cells would effectively get infected), we find that an MOI between once and twice the number of genome segments yields that efficiency, even for multipartite viruses with many segments. This prediction coincides with the MOI values that are indeed found experimentally in plants [8]. Contrary to that, infections in animals are frequently subject to bottlenecks, events for which the MOI becomes severely reduced. The onset of viral infections in animals, as well as intrahost dispersal, are processes that involve a very small number of viral particles [32,33], too small for a multipartite virus to achieve efficient infection. Hence, the asymmetry in the host distribution of multipartite viruses may derive from differences in the characteristic MOIs for plants and animals, which in fact are a consequence of the physiological constraints governing viral transmission and dispersal in different organisms.

(b) *On the origin of multipartite viruses*
Should the evolution of multipartite viruses in nature proceed through competition among segments with different lengths, then a relatively small number of segments is expected in the light of our model. This may have been the case for multipartite viruses belonging to the families Geminiviridae, Secoviridae and Bromoviridae (the former two composed of two segments, and the latter of three segments). Among them, the family Geminiviridae is interesting, as it contains bipartite genera as well as non-segmented ones [10].

However, some of the multipartite viruses found in nature present a much larger number of segments. In particular, members of the family *Nanoviridae* are composed of six or eight segments [34], so that the MOI that is required to get them fixed becomes of the order of 100 (or even higher). This value is very unlikely to be attained in nature. Therefore, other conceptual frameworks are needed in order to explain the origin of these highly multipartite viruses. Two alternative hypotheses can be proposed in this respect. The first is that viral capsids and genome size have co-evolved, in such a way that as the genome becomes segmented a smaller capsid is recruited. As the stability of the viral particle depends on the chemical interaction between genome and capsid, it can be expected that if the new capsid fits the size of the genome segments, the relative fitness advantage of a further segmentation would increase. An argument supporting this idea is the fact that viral capsids in nanovirus indeed fit the segment size; so no multiple-segment hypothetical progenitor could be packed into them. A second hypothesis consists of accepting that there has been only one (or maybe two) true segmentation event, favourably selected by a moderate MOI, and the rest of segments have been recruited as genes captured from other viruses. In this respect, we recall that interspecific recombination and gene transfer events are widespread in multipartite viruses [35,36] and are thought to have played a role in the particular evolution of nanoviruses [37,38].

REFERENCES
1 Eigen, M. 1993 Viral quasispecies. *Scient. Am.* **296**, 32–39.
2 Roosinck, M. J. 1997 Mechanisms of plant virus evolution. *Annu. Rev. Phytopathol.* **35**, 191–209. (doi:10.1146/annurev.phyto.35.1.191)
3 Manrubia, S. C. & Lázaro, E. 2006 Viral evolution. *Phys. Life Rev.* **3**, 65–92. (doi:10.1016/j.plrev.2005.11.002)
4 Szathmary, E. 1992 Natural selection and dynamical coexistence of defective and complementing virus segments. *J. Theoret. Biol.* **157**, 383–406. (doi:10.1016/S0022-5193(05)80617-4)
5 Turner, P. E. & Chao, L. 1998 Sex and the evolution of intrahost competition in RNA virus ϕ6. *Genetics* **150**, 523–532.
6 Ojosnegros, S., García-Arriaza, J., Escarmís, C., Manrubia, S. C., Perales, C., Arias, A., García Mateu, M. & Domingo, E. 2011 Viral genome segmentation can result from a trade-off between genetic content and particle stability. *PLoS Genet.* **7**, e1001344. (doi:10.1371/journal.pgen.1001344)
7 Palukaitis, P. & García-Arenal, F. 2003 Cucumoviruses. *Adv. Virus Res.* **62**, 241–323. (doi:10.1016/S0065-3527(03)62005-1)
8 Betancourt, M., Fereres, A., Fraile, A. & García-Arenal, F. 2008 Estimation of the effective number of founders that initiate an infection after aphid transmission of a multipartite plant virus. *J. Virol.* **82**, 12 416–12 421. (doi:10.1128/JVI.01542-08)
9 González-Jara, P., Fraile, A., Canto, T. & García-Arenal, F. 2009 The multiplicity of infection of a plant virus varies during colonization of its eukaryotic host. *J. Virol.* **83**, 7487–7494. (doi:10.1128/JVI.00636-09)
10 Lazarowitz, S. D. 2007 Plant viruses. In *Fields virology*, 5th edn., vol. 1 (eds D. M. Knippe & P. M. Howley), pp. 641–706. Philadelphia, PA: Lippincott Williams and Wilkins.
11 Nee, S. 1987 The evolution of multicompartmental genomes in viruses. *J. Mol. Evol.* **25**, 277–281. (doi:10.1007/BF02603110)
12 Holland, J. J. 1990 *Virology*, vol. 1 (eds B. N. Fields & D. M. Knipe), pp. 151–165. New York, NY: Raven.
13 Bangham, C. R. M. & Kirkwood, T. B. L. 1990 Defective interfering particles: effects in modulationg virus growth and persistence. *Virology* **179**, 821–826. (doi:10.1016/0042-6822(90)90150-P)
14 Roux, L., Simon, A. E. & Holland, J. J. 1991 Effects of defective interfering viruses on virus replication and pathogenesis *in vitro* and *in vivo*. *Adv. Virus Res.* **40**, 181–211. (doi:10.1016/S0065-3527(08)60279-1)
15 Kirkwood, T. B. L. & Bangham, C. R. M. 1994 Cycles, chaos, and evolution in virus cultures: a model of defective interfering particles. *Proc. Natl Acad. Sci. USA* **91**, 8685–8689. (doi:10.1073/pnas.91.18.8685)
16 Frank, S. A. 2000 Within-host spatial dynamics of viruses and defective interfering particles. *J. Theoret. Biol.* **206**, 279–290. (doi:10.1006/jtbi.2000.2120)
17 Szathmary, E. 1993 Co-operation and defection: playing the field in virus dynamics. *J. Theoret. Biol.* **165**, 341–356. (doi:10.1006/jtbi.1993.1193)
18 Wilke, C. O. & Novella, I. S. 2003 Phenotypic mixing and hiding may contribute to memory in viral quasispecies. *BMC Microbiol.* **3**, 11. (doi:10.1186/1471-2180-3-11)

19 Chao, L. 1991 Levels of selection, evolution of sex in RNA viruses, and the origin of life. *J. Theoret. Biol.* **153**, 229–246. (doi:10.1016/S0022-5193(05)80424-2)

20 García-Arriaza, J., Manrubia, S. C., Toja, M., Domingo, E. & Escarmís, C. 2004 Evolutionary transition toward defective RNAs that are infectious by complementation. *J. Virol.* **78**, 11 678–11 685. (doi:10.1128/JVI.78.21.11678-11685.2004)

21 O'Neill, F. J., Maryon, E. B. & Carroll, D. 1982 Isolation and characterization of defective simian virus 40 genomes which complement for infectivity. *J. Virol.* **43**, 18–25.

22 Geigenmüller-Gnirke, U., Weiss, B., Wright, R. & Schlesinger, S. 1991 Complementation between Sindbis viral RNAs produces infectious particles with a bipartite genome. *Proc. Natl Acad. Sci. USA* **88**, 3253–3257.

23 Kim, K. H., Narayanan, K. & Makino, S. 1997 Assembled coronavirus from complementation of two defective interfering RNAs. *J. Virol.* **71**, 3922–3931.

24 García-Arriaza, J., Ojosnegros, S., Dávila, M., Domingo, E. & Escarmís, C. 2006 Dynamics of mutation and recombination in a replicating population of complementing, defective viral genomes. *J. Mol. Biol.* **360**, 558–572. (doi:10.1016/j.jmb.2006.05.027)

25 Smouse, P. E. 1976 The implications of density-dependent population growth for frequency and density-dependent selection. *Am. Nat.* **110**, 849–860. (doi:10.1086/283107)

26 Novella, I. S., Reissig, D. D. & Wilke, C. O. 2004 Density-dependent selection in vesicular stomatitis virus. *J. Virol.* **78**, 5799–5804. (doi:10.1128/JVI.78.11.5799-5804.2004)

27 Ball, L. A. 2007 Virus replication strategies. In *Fields virology*, 5th edn., vol. 1 (eds D. M. Knippe & P. M. Howley), pp. 119–140. Philadelphia, PA: Lippincott Williams and Wilkins.

28 Belsham, G. J. 2005 Translation and replication of FMDV RNA. *Curr. Top. Microbiol. Immunol.* **288**, 43–70. (doi:10.1007/3-540-27109-0_3)

29 Grande-Pérez, A., Lázaro, E., Lowenstein, P., Domingo, E. & Manrubia, S. C. 2005 Suppression of viral infectivity through lethal defection. *Proc. Natl Acad. Sci. USA* **102**, 4448–4452. (doi:10.1073/pnas.0408871102)

30 Iranzo, J. & Manrubia, S. C. 2009 Stochastic extinction of viral infectivity through the action of defectors. *Europhys. Lett.* **85**, 18001. (doi:10.1209/0295-5075/85/18001)

31 Gutiérrez, S., Yvon, M., Thébaud, G., Monsion, B., Michalakis, Y. & Blanc, S. 2010 Dynamics of the multiplicity of cellular infection in a plant virus. *PLoS Pathog.* **6**, e1001113. (doi:10.1371/journal.ppat.1001113)

32 Frost, S. D. W., Dumaurier, M. J., Wain-Hobson, S. & Leigh-Brown, A. J. 2001 Genetic drift and within host metapopulation dynamics of HIV-1 infection. *Proc. Natl Acad. Sci. USA* **98**, 6975–6980. (doi:10.1073/pnas.131056998)

33 Kuss, S. K., Etheredge, C. A. & Pfeiffer, J. K. 2008 Multiple host barriers restrict poliovirus trafficking in mice. *PLoS Pathog.* **4**, e1000082. (doi:10.1371/journal.ppat.1000082)

34 Gronenborn, B. 2004 Nanoviruses: genome organisation and protein function. *Vet. Microbiol.* **98**, 103–109. (doi:10.1016/j.vetmic.2003.10.015)

35 Chare, E. R. & Holmes, E. C. 2006 A phylogenetic survey of recombination frequency in plant RNA viruses. *Arch. Virol.* **151**, 933–946. (doi:10.1007/s00705-005-0675-x)

36 Lefeuvre, P., Lett, J. M., Varsani, A. & Martin, D. P. 2009 Widely conserved recombination patterns among single-stranded DNA viruses. *J. Virol.* **83**, 2697–2707. (doi:10.1128/JVI.02152-08)

37 Gibbs, M. J. & Weiller, G. F. 1999 Evidence that a plant virus switched hosts to infect a vertebrate and then recombined with a vertebrate-infecting virus. *Proc. Natl Acad. Sci. USA* **96**, 8022–8027. (doi:10.1073/pnas.96.14.8022)

38 Hu, J. M., Fu, H. C., Lin, C. H., Su, H. J. & Yeh, H. H. 2007 Reassortment and concerted evolution in banana bunchy top virus genomes. *J. Virol.* **81**, 1746–1761. (doi:10.1128/JVI.01390-06)