The relationship between mutation frequency and replication strategy in positive-sense single-stranded RNA viruses

Gaël Thébaud¹, Joël Chadœuf², Marco J. Morelli³, John W. McCauley⁴ and Daniel T. Haydon⁴,*

¹Institut National de la Recherche Agronomique (INRA), UMR BGPI, Cirad TA A-54/K, Campus de Baillarguet, 34398 Montpellier cedex 5, France
²INRA, UR546 Biostatistique et Processus Spatiaux, Domaine Saint-Paul, 84914 Avignon, France
³Boyd Orr Centre for Population and Ecosystem Health, Faculty of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK
⁴The National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

For positive-sense single-stranded RNA virus genomes, there is a trade-off between the mutually exclusive tasks of transcription, translation and encapsidation. The replication strategy that maximizes the intracellular growth rate of the virus requires iterative genome transcription from positive to negative, and back to positive sense. However, RNA viruses experience high mutation rates, and the proportion of genomes with lethal mutations increases with the number of replication cycles. Thus, intracellular mutant frequency will depend on the replication strategy. Introducing apparently realistic mutation rates into a model of viral replication demonstrates that strategies that maximize viral growth rate could result in an average of 26 mutations per genome by the time plausible numbers of positive strands have been generated, and that virus viability could be as low as 0.1 per cent. At high mutation rates or when a high proportion of mutations are deleterious, the optimal strategy shifts towards synthesizing more negative strands per positive strand, and in extremis towards a ‘stamping-machine’ replication mode where all the encapsidated genomes come from only two transcriptional steps. We conclude that if viral mutation rates are as high as current estimates suggest, either mutation frequency must be considerably higher than generally anticipated and the proportion of viable viruses produced extremely small, or replication strategies cannot be optimized to maximize viral growth rate. Mechanistic models linking mutation frequency to replication mechanisms coupled with data generated through new deep-sequencing technologies could play an important role in improving the estimates of viral mutation rate.

Keywords: analytical model; generation interval; individual-based model; Poliovirus; polyprotein; population dynamics

1. INTRODUCTION

Life-history theory is a cornerstone of evolutionary biology. It mainly relies on predicting how life histories may evolve to maximize individual life-time reproductive success or intrinsic growth rate (Stearns 1992). Most studies focus on the life-history theory of multi-cellular organisms. However, the reproductive biology of microbial organisms is in some respects much simpler and more amenable to detailed understanding. In addition, some biological characteristics of viruses make them particularly well suited to the application of optimality theory: their enormous population size, high mutation rate and short generation time provide a lot of variation in life-history traits on which selection can act efficiently. Here we consider optimal replication strategies for positive-sense single-stranded RNA (ssRNA(þ)) viruses within a cell.

The ssRNA(þ) viruses are those in which the viral particle (virion) contains the sense strand that can be translated directly into protein. For their replication within cells, ssRNA(þ) viruses must use their genetic material for at least three mutually exclusive and sequential activities. The first is translation, the production from RNA of essential viral proteins that are required for molecular replication machinery and the construction of capsids. The second is transcription, required to amplify the number of RNA strands available for further translation and/or subsequent encapsidation activities. The third is encapsidation whereby genomes are packaged into virus capsids, producing completed virions that exclude the genome from the other two processes.

Viral replication can only occur within a cell. When a single ssRNA(þ) virus enters a host cell, it first undertakes translation through the use of the cellular machinery, which produces structural proteins as well as RNA-dependent RNA polymerase and any other viral proteins required for transcription (we will refer to this protein complex as ‘replicase’). If the virus is to increase its rate of protein production, it will need to generate more positive strands for translation, and these are produced through successive rounds of transcription whereby negative strands are copied from positive strands,
and positive strands then copied from negative strands. As each generation of transcription occurs (negative copied from positive, and positive copied from negative), the population of positive strands increases. These positive strands can be used for (i) translation and the production of more replicase and structural proteins required for capsid assembly, (ii) further transcription and production of negative strands, or (iii) encapsidation and production of virions. At some point, the extent of viral replication within the cell causes the release of virions that go on to infect additional cells.

The optimal balance of these activities will depend on how virus fitness is defined. At least in their early stages, viral infections are characterized by a rapidly increasing viral load within an individual. It is therefore a reasonable assumption that viral fitness will be increased by rapid production of daughter virions, and that the virus should maximize its Malthusian fitness (Fisher 1930), i.e. the exponential growth rate virion production within the cell (Krakauer & Komarova 2003). Regoes et al. (2005) have addressed the question of the optimal replication strategy by examining with a simple analytical model the balance of transcription and translation that leads to maximizing the growth rate of the population of positive strands. However, their study does not consider the effects of encapsidation or mutation on this optimal strategy (figure 1). Encapsidation is required for the production of virions but it prevents further replication from the encapsidated positive strands; thus, it may play an important role in defining the optimal replication strategy.

The optimal replication strategy may also be influenced by the mutation rate, which is variously cited to be between $10^{-5}$ and $2 \times 10^{-3}$ mutations per nucleotide per replication event (mut/nt/rep) among ssRNA(+) viruses (Holland et al. 1982; Drake 1993; Duffy et al. 2008). The number of mutations (many of which are deleterious) will increase through successive generations of transcription, and this may favour a replication strategy that increases the number of genomes produced from early generation transcripts (Chao et al. 2002; Drake 2007; Duffy et al. 2008). Because the mutation load accumulates with the number of transcriptional generations, the inclusion of mutation dynamics in an exact model of viral replication requires an estimate of the contribution of each transcriptional generation to the positive strand population, and assumptions about how mutation load is related to viral viability. Here, we present alternative models for viral replication within cells that include encapsidation and that incorporate the demographic contribution from different generations so that mutation dynamics can be studied.

Our analyses are based on processes typical of Picornaviridae and Potyviridae, two groups of ssRNA(+) viruses that have their genome translated into a single self-cleavable polyprotein. The Picornaviridae comprise many important animal and human viruses (e.g. Poliovirus, Hepatitis A virus, Foot-and-mouth disease virus) and the Potyviridae include many viruses infecting staple or commercial crops (e.g. Potato virus Y, Plum pox virus, Yam mosaic virus). More specifically, we chose parameters applying to Poliovirus whose molecular biology is the best understood. We address two main questions: (i) How is the optimal replication strategy affected by encapsidation and by the presence of deleterious mutation? (ii) What is the expected number of mutations in virus genomes that results from these optimal replication strategies? Our models were also used to address a series of secondary questions that include the changes through time in the ratio of positive to negative strands (or replicate) and the proportion of viable virions for different mutation rates.

2. MATERIAL AND METHODS

(a) Virus intracellular growth

To model virus replication, we draw heavily on the notation, parameter values and processes described by Regoes et al. (2005). All parameters used in the models described below are summarized in table 1. The within-cell treatment of positive strands assumes no complementation and is represented by the following sequence of events: (i) translation into $n_1$ self-cleavable polyproteins, each providing all the viral proteins, including one replicase and one capsid protomer (a full capsid comprises $n_L$ protomers, so each positive strand generates $n_L/n_1$ full capsids), (ii) encapsidation of a proportion $n_L/n_1$ of the positive strands, and (iii) transcription of the remaining free genomes into $n_2$ negative strands. As soon as a negative strand is completed, it is itself subject to transcription, which generates $n_2$ positive strands, and the process starts again. A given replication strategy is defined here by a set of values for the triplet $(r_{f0}, r_p, r_1)$.

More specifically, during the first phase (translation), the cellular machinery is assumed to be non-limiting: the first ribosome moves from the 5'- to the 3'-end of the positive-sense RNA, followed by $r_1$ other evenly spaced ribosomes (with a limit of $n_L$ ribosomes fitting simultaneously on the same RNA strand). The time required by a ribosome to move along a whole positive strand and produce a single polyprotein is $T_L$, thus, translation of the first polyprotein from a given positive strand is always completed after a delay $\delta = n_L$, and each of the subsequent $n_1 - 1$ polyproteins is fully translated after an additional delay $e_L = n_L/n_2$. As a result, the time required for translation of the $n_L$ polyproteins is $T_L = n_L(1 + (n_L - 1)/n_L)$. The second phase (encapsidation) is assumed to happen instantaneously after the $n_L$ polyproteins have been translated from the positive strand, and precludes any further replication of the encapsidated positive strands (adding a modest delay prior to encapsidation should not change the relative fitness of different
replication strategies). During the third phase (transcription), the maximum number of replicases per template strand is \( n_x \) and the time taken to produce a single RNA strand is \( \tau_c \); thus, the first negative strand is fully transcribed from a given positive strand after a delay \( \delta_N = T_L + \tau_c \), transcription of each subsequent negative strand is completed after an additional delay \( \delta_N \). Similarly, the first positive strand is fully transcribed from a given negative strand after a delay \( \delta_p \) and transcription of each subsequent positive strand is completed after an additional delay \( \delta_p \) (figure 2). When viral replicase is not limiting, \( \delta_p = \tau_c \), and \( \delta_N = \delta_p = \tau_c / n_x \). There are three different ways (non-mutually exclusive) in which replicase can be a limiting factor and thus delay replication: (i) when \( n_x < n_{\alpha} \), a suboptimal number of replicases is fitted simultaneously on each positive strand template, which increases the delay between the synthesis of successive positive strands from \( \delta_N = \tau_c / n_x \) to \( \delta_N = \tau_c / n_{\alpha} \), (ii) when \( n_x < n_{\alpha} \), not limiting, \( \delta_N = \delta_p \), and the time taken to produce a single RNA strand is \( \tau_c \); thus, the first negative strand is fully transcribed from a given positive strand after a delay \( \delta_N \), and the time between successive negative strands is \( \delta_N \). Each negative strand is in turn immediately used as a template for the transcription of \( n_x \) new positive strands; the first positive strand appears after a delay \( \delta_p \) and the time between successive positive strands is \( \delta_p \). Here, \( R_0 = n_x \delta_p = 8 \).

**Figure 2.** Transcriptional events involved in positive-sense single-stranded RNA virus replication. A given positive strand (in black) is first translated into proteins, and then used as a template for the transcription of \( n_x \) negative strands (in grey); the first negative strand appears after a delay \( \delta_N \) and the time between successive negative strands is \( \delta_N \). Each negative strand is in turn immediately used as a template for the transcription of \( n_x \) new positive strands; the first positive strand appears after a delay \( \delta_p \) and the time between successive positive strands is \( \delta_p \).

**Table 1.** Parameters used in the models.

<table>
<thead>
<tr>
<th>parameter</th>
<th>interpretation</th>
<th>default value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_N )</td>
<td>number of negative strands transcribed from each positive strand</td>
<td>to be estimated</td>
</tr>
<tr>
<td>( \tau_P )</td>
<td>number of positive strands transcribed from each negative strand</td>
<td>to be estimated</td>
</tr>
<tr>
<td>( \tau_L )</td>
<td>number of polyproteins translated from each positive strand</td>
<td>to be estimated</td>
</tr>
<tr>
<td>( n_x )</td>
<td>maximum number of replicases per RNA strand</td>
<td>6.5</td>
</tr>
<tr>
<td>( \tau_c )</td>
<td>time taken to transcribe one RNA strand</td>
<td>1.2 min</td>
</tr>
<tr>
<td>( \delta_N )</td>
<td>delay prior to production of first negative strand</td>
<td>( T_L + \tau_c )</td>
</tr>
<tr>
<td>( \delta_P )</td>
<td>delay prior to production of first positive strand</td>
<td>see electronic supplementary material, appendix 1</td>
</tr>
<tr>
<td>( e_N )</td>
<td>interval between transcription of successive negative strands</td>
<td>( \tau_c / \min(n_{\alpha}, n_L) )</td>
</tr>
<tr>
<td>( e_P )</td>
<td>interval between transcription of successive positive strands</td>
<td>( \tau_c / \min(n_{\alpha}, n_L) )</td>
</tr>
<tr>
<td>( n_L )</td>
<td>maximum number of ribosomes per positive strand</td>
<td>30</td>
</tr>
<tr>
<td>( \tau_L )</td>
<td>time taken to translate one positive strand into one polyprotein</td>
<td>6.25 min</td>
</tr>
<tr>
<td>( \delta_L )</td>
<td>delay prior to production of first polyprotein</td>
<td>( n_L / n_{\alpha} )</td>
</tr>
<tr>
<td>( t )</td>
<td>delay between completion of successive polyproteins</td>
<td>( n_L (1 + (n_L - 1) / n_{\alpha}) )</td>
</tr>
<tr>
<td>( g )</td>
<td>number of generations of positive strands</td>
<td>60</td>
</tr>
<tr>
<td>( \delta_t )</td>
<td>number of generations of positive strands</td>
<td>—</td>
</tr>
<tr>
<td>( \mu )</td>
<td>mutation rate (per nucleotide per replication event)</td>
<td>( 4.5 \times 10^{-4} )</td>
</tr>
<tr>
<td>( S )</td>
<td>genome size (in nucleotides)</td>
<td>7500</td>
</tr>
<tr>
<td>( f_0 )</td>
<td>proportion of mutations that are lethal</td>
<td>0.4</td>
</tr>
<tr>
<td>( R_0 )</td>
<td>total progeny number of a single positive strand</td>
<td>( n_x \mu (1 - n_x / n_{\alpha}) )</td>
</tr>
<tr>
<td>( r )</td>
<td>intrinsic instantaneous growth rate</td>
<td>to be maximized</td>
</tr>
<tr>
<td>( y_i )</td>
<td>proportion of the progeny of one positive strand within each of ( n ) histogram categories of width ( w = e_P = e_N )</td>
<td>see electronic supplementary material, appendix 2</td>
</tr>
<tr>
<td>( m )</td>
<td>mutation-induced mortality rate per time unit</td>
<td>—</td>
</tr>
<tr>
<td>( T_g )</td>
<td>mean intergeneration interval</td>
<td>—</td>
</tr>
</tbody>
</table>

**Proc. R. Soc. B (2010)**
between synthesis of each positive strand and synthesis of the corresponding polyproteins (and then, replicate). As before, parameter values were taken from the literature as described by Regoes et al. (2005) (time to translate one polyprotein, \( \tau_x = 6.25 \) min; time to transcribe one RNA strand, \( \tau_y = 1.2 \) min; maximum number of ribosomes per RNA strand, \( n_t = 10 \); maximum number of replicates per RNA strand, \( n_t = 6.5 \)). Viral capsids are highly symmetric in their construction, and their assembly requires multiple copies of structural proteins; for Poliovirus, the full capsid comprises \( n_c = 60 \) protomers (Minor et al. 1986).

In order to compute the number of viable virions through time, the proportion of virions with a genome affected by at least one lethal mutation must be estimated. A full genome comprises \( S \) nucleotides, each of which is synthesised with a mutation rate \( \mu \); a proportion \( f_m \) of these mutations are assumed to be lethal. Thus, the random variable representing the number of lethal mutations incurred by a given genome at the \( g \)th generation of positive strands, \( X_g \), can be modelled by a Poisson distribution with parameter \( \lambda_g = 2gS\mu \) (the factor 2 comes from the two replication events occurring between successive generations of positive strands). The probability that a genome is viable corresponds to the null class of this distribution: \( P(X_g = 0) = e^{-\lambda_g} \). The effect of mutation was included in the A-IBM as an additional step where the total number of virions produced at generation \( g \) was multiplied by \( e^{-\lambda_g} \) before merging the distributions corresponding to the different generations. The complete sequence of Poliovirus genome being about 7.5 kb long (Kitamura et al. 1981; Racaniello & Baltimore 1981), we took \( S = 7500 \) nucleotides. Based on previous empirical work on other RNA viruses, \( f_m \) was taken to be 0.4 (Sanjuán et al. 2004; Carrasco et al. 2007), while the other mutations were assumed to be neutral. Mutation rates are notoriously difficult to estimate for a variety of reasons (e.g. box 2 in Duffy et al. 2008); thus, in our analyses, the mutation rate characterizing Poliovirus replicase was included as a variable ranging from \( 10^{-4} \) to \( 1.1 \times 10^{-3} \) mut/nt/rep with a default value of \( 4.5 \times 10^{-4} \) mut/nt/rep, measured after just two rounds of replication with the Poliovirus polymerase (Rodriguez-Wells et al. 2001). The asymptotic growth rate of the A-IBM, estimated based on a log-linear regression, was used as a benchmark for further estimations of virion growth with an analytical model that ensures shorter computing times and exact asymptotic growth rates.

(c) Analytical model for virion growth rate

To build the analytical model, we define the total progeny number \((R_0)\) of a precursor positive strand as the number of positive strands synthesized from negative strands transcribed directly from the precursor strand and that are available for future transcription. The distribution of the times at which individual progeny are synthesized defines the generation interval distribution (see figure 2, for an example with \( R_0 = 8 \) corresponding to a specific replication strategy (as defined by a set of values for \( r_{so}, \gamma_{so}, r_{p} \) and \( r_{L} \)).

There is a mathematical relationship between \( R_0 \), the growth rate \( r \) and the generation interval distribution (Wallinga & Lipsitch 2007). For the population of positive strands, any replication strategy defines \( R_0 \) and the generation interval distribution, which can be represented by a histogram with equal category width \( \varepsilon \) (time between two consecutive strands in the progeny). As a result, \( r \) can be derived numerically through rearranging equation (3.6) in Wallinga & Lipsitch (2007) as:

\[
R_0 = re^\epsilon \sum_{i=1}^{n} (y_i e^{-\epsilon}) ,
\]

where \( y_i \) is the proportion of the progeny of one positive strand within each of \( n \) histogram categories of width \( \epsilon \), and \( a_0 \) being the first bound of the histogram starting at \( \varepsilon /2 \) before the first progeny strand. When encapsidation is not taken into account, \( R_0 = r_{so} r_{L} \) with encapsidation, \( R_0 = r_{so} r_{L} / (1 - n_c n_t) \). This relationship between \( r \) and \( R_0 \) enabled efficient identification of the replication strategy corresponding to the optimal virion growth rate in the absence of mutation.

In the presence of mutation, the instantaneous growth rate of viable virions, \( r_v \), is negatively affected by a lethal mutation rate \( m \). Thus, the growth rate of viable virions is \( r_v = r - m \approx r - \lambda T_g \) where \( \lambda = 2S\mu \) is the expected number of lethal mutations incurred by a progeny strand in one generation, and \( T_g \) is the mean inter-generation interval. The growth rate in the presence of mutation, \( r_v \), can either be approximated from our exact analytical model for \( r_v \) in which case

\[
r_v \approx r - \frac{\lambda}{T_g}, \tag{2.1}
\]

or from the classical approximation

\[
r_v \approx \ln(r_0)/T_g \text{ in which case}
\]

\[
r_v \approx \ln(r_0) - \frac{\lambda}{T_g}. \tag{2.2}
\]

In both approximations, the rescaling factor is known to be underestimated by \( T_g \) (Wallinga & Lipsitch 2007). Thus, the first approximation overestimates \( r_v \) and the second one is an underestimation: we used both approximations in combination as a way to bracket \( r_v \), the growth rate of the viable virions. Note that in the absence of mutation (\( \lambda = 0 \)), equation (2.1) gives the exact growth rate, while equation (2.2) corresponds to a classical approximation (Begon et al. 1990; Case 1999; Regoes et al. 2005).

(d) Optimization procedure

In order to find the optimal set of parameter values (for \( r_{so}, r_{p}, r_{L} \)) that maximizes the Malthusian fitness (i.e. population growth rate), we used the simple optimization procedure consisting of (i) filling a cube with growth rates computed for each combination of the integer values of the three parameters (\( r_{so}, r_{p} \) and \( r_{L} \)) and (ii) finding the coordinates of the maximum growth rate. After a few initial heuristic searches, \( r_{so} \) was varied between 1 and 130, \( r_{p} \) between 1 and 100 and \( r_{L} \) between 1 and 30. We checked that the growth rate varied smoothly over the parameter space and that no other parameter than \( r_{p} \) (see below) was at its optimum on the boundary of the explored parameter space.

3. RESULTS

(a) Optimal growth rate for positive strands and virions in the absence of mutation

The analytical models enabled defining the set of parameters corresponding to the optimal replication strategy (figure 3). The maximum growth rate of positive strands (\( r_v = 0.317 \)) corresponds to \( r_{so} = 2, r_{p} = 100 \) and \( r_{L} = 13 \) (black curves). For virions, the maximum growth rate (\( r_v = 0.3 \)) is obtained for \( r_{so} = 2, r_{p} = 100 \) and \( r_{L} = 12 \) (hereafter, these values define the default replication strategy). Thus, encapsidation of \( r_{L}/n_t = 20 \) per cent of the progeny genomes only reduces the growth rate by 5.4 per cent and has almost no effect on the optimal set of parameters (dark grey curves).

growth rate but has little impact on the optimal set of parameters, except that $r_N$ reaches higher values for both approximations (figure 3b). With increasing mutation rates, the optimal strategy shifts towards synthesizing more negative strands per positive strands and, using equation (2.1), more polyprotein (table 2). Thus, more strands of the first generation can be encapsidated, which is favoured because most of the virions from the next generations are non-viable. For the highest mutation rates, this shift leads to an optimal replication strategy that is in effect similar to a ‘stamping-machine’ replication mode (Chao et al. 2002; Duffy et al. 2008) with a very small growth rate and all the progeny genomes coming from only two transcriptional steps.

### (c) Distribution of mutations in the population of virions

(i) Changes in the proportion of viable virions

Virion yield increases exponentially with the number of generations (figure 4a, dashed line), but the genetic quality of the encapsidated genomes decreases exponentially at the same time. This trade-off affects the growth rate of the population of viable virions (table 2, figure 3). It also leads to an exponential decrease in the proportion of viable virions with increasing generation number, at a rate that depends on the assumed mutation rate, as shown using the A-IBM. For $\mu = 10^{-5}$ mut/nt/rep, the viability of most genomes is preserved, while for $\mu = 10^{-3}$ mut/nt/rep, the few viable virions are overwhelmed by a rapidly increasing number of non-viable virions; our default mutation rate ($\mu = 4.5 \times 10^{-4}$ mut/nt/rep) corresponds to a decline in the proportion of viable virions from 1 to as little as $3 \times 10^{-6}$ within less than five generations (figure 4a).

(ii) Frequency of mutants in the viral yield from a single cell

The peak number of positive strands within one Poliovirus-infected cell has been estimated to be approximately 76 000 (Novak & Kirkegaard 1991). By the time this value is reached in the A-IBM, each of the 15 320 encapsidated genomes has on average 26 mutations for $\mu = 4.5 \times 10^{-4}$ mut/nt/rep (0.11% of them are viable), while mutation rates of $10^{-3}$ and $10^{-2}$ mut/nt/rep result in an average of 5.8 and 59 mutations per genome, respectively (figure 4b). However, a high mutation rate still produces several hundreds of viable virions, as illustrated by the A-IBM (figure 5a).

### (d) Ratio of viral molecules

(i) Ratio of free positive strands to negative strands

The A-IBM shows that the initial steps of viral replication should be dominated by transcriptional generations that overlap little because of the sequential nature of the process (figure 5a), which leads to wide oscillations in the ratio of positive to negative strands (figure 5b, black curve). For the default replication strategy, this ratio is expected to vary between 1 and 40 and to stabilize around 9.6 after some time.

(ii) Ratio of free positive strands to virions

Under the assumption that all the available capsid protein is used immediately, we could derive a straightforward expression for the ratio of positive strands to virions.

---

**Figure 3.** Optimal growth rate conditional on the value of each parameter defining the replication strategy. (a) $r_P$, (b) $r_N$, and (c) $r_L$. Growth rates of positive strands (black), virions (dark grey) and viable virions for $\mu = 4.5 \times 10^{-4}$ mut/nt/rep (light grey for both the overestimation and underestimation) are obtained with the analytical model. Dashed lines correspond to the coordinates of the optimum, except for (a) where they correspond to 99% of the optimal growth rate. The number of negative strands, positive strands and polyproteins synthesized from one template strand are represented by $r_N$, $r_P$ and $r_L$, respectively.

In both cases, the optimal value for $r_P$ is at the edge of the explored parameter space, but any increase in $r_P$ leads to a very small increase in $r$ (data not shown), so $r_P$ has in fact no finite optimal value. However, the growth rates of positive strands and virions increase asymptotically with $r_P$ (e.g. $r$ increases by less than $10^{-8}$ when $r_P$ increases from 500 to 1000), and they reach 99 per cent of their maximum value (taken at $r_P = 1000$) for $r_P = 35$ and $r_P = 30$, respectively (figure 3a, dashed lines). Thus, the optimal set of parameter values corresponds to a replication strategy in which each positive strand is transcribed into only two negative strands, each negative strand is then transcribed into as many positive strands as possible and translation is adjusted accordingly: more polyprotein synthesis (higher $r_L$) requires too much time, while less polyprotein synthesis depletes the pool of replicase, which hinders replication.

**(b) Optimal growth rate in the presence of mutation**

Integrating mutation ($\mu = 4.5 \times 10^{-4}$ mut/nt/rep) into the analytical model significantly reduces the optimal
examine how mutation accumulation is likely to affect
the overall viral yield, it becomes straightforward to
replication strategies in two ways: first, they explicitly
These new models advance our understanding of virus
growth of positive-sense single-stranded RNA viruses.
We have developed analytical and computational models

4. DISCUSSION
We have developed analytical and computational models
of within-cell replication during the phase of exponential
growth of positive-sense single-stranded RNA viruses.
These new models advance our understanding of virus
replication strategies in two ways: first, they explicitly
include the encapsidation process; and second, because
they enable the estimation of the demographic contribution of
different transcriptional generations to
the overall viral yield, it becomes straightforward to
examine how mutation accumulation is likely to affect
the number of viable virions produced from a single
infected cell.

Table 2. Effect of different mutation rates \((\text{mut/nt/rep})\) on
the parameters corresponding to the maximum growth rate
\(r_v\) of the viable virions (bracketed by equations (2.1) and
(2.2) that provide overestimated and underestimated values
of \(r_v\), respectively). The number of negative strands, positive
strands and polyproteins synthesized from one template
strand are represented by \(r_N\), \(r_P\) and \(r_L\), respectively.

<table>
<thead>
<tr>
<th>mutation rate</th>
<th>equation (2.1)</th>
<th>equation (2.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(r_v)</td>
<td>(r_N)</td>
</tr>
<tr>
<td>0</td>
<td>0.30011 2</td>
<td>100 12</td>
</tr>
<tr>
<td>(1 \times 10^{-6})</td>
<td>0.29976 2</td>
<td>100 12</td>
</tr>
<tr>
<td>(1 \times 10^{-5})</td>
<td>0.29657 2</td>
<td>100 12</td>
</tr>
<tr>
<td>(5 \times 10^{-5})</td>
<td>0.28314 4</td>
<td>100 11</td>
</tr>
<tr>
<td>(1 \times 10^{-4})</td>
<td>0.26843 7</td>
<td>100 11</td>
</tr>
<tr>
<td>(2 \times 10^{-4})</td>
<td>0.24205 10</td>
<td>100 10</td>
</tr>
<tr>
<td>(3 \times 10^{-4})</td>
<td>0.21605 11</td>
<td>100 11</td>
</tr>
<tr>
<td>(4 \times 10^{-4})</td>
<td>0.19027 12</td>
<td>100 12</td>
</tr>
<tr>
<td>(4.5 \times 10^{-4})</td>
<td>0.17749 12</td>
<td>100 12</td>
</tr>
<tr>
<td>(5 \times 10^{-4})</td>
<td>0.16574 22</td>
<td>100 11</td>
</tr>
<tr>
<td>(6 \times 10^{-4})</td>
<td>0.14368 24</td>
<td>100 12</td>
</tr>
<tr>
<td>(7 \times 10^{-4})</td>
<td>0.12246 33</td>
<td>100 11</td>
</tr>
<tr>
<td>(8 \times 10^{-4})</td>
<td>0.10306 48</td>
<td>100 12</td>
</tr>
<tr>
<td>(9 \times 10^{-4})</td>
<td>0.08538 60</td>
<td>100 12</td>
</tr>
<tr>
<td>(1 \times 10^{-3})</td>
<td>0.06946 78</td>
<td>100 13</td>
</tr>
<tr>
<td>(1.1 \times 10^{-3})</td>
<td>0.05533 120</td>
<td>100 15</td>
</tr>
</tbody>
</table>

For one synthesized positive strand, \(1 - r_L/n_L\) remains
free and \(n_L/n_L - 1 = 4\) (figure 5a).

(iii) Ratio of free positive strands to replicase
If we used a similar calculation, the ratio of free positive
strands to replicase should be: \(n_L/n_L = 6.7 \times 10^{-2}\). However, this analytical result corresponds to the
critical time when replicase may be limiting; thus, it is
expected to differ from the average ratio over the whole
replication cycle. The A-IBM was used to track the
amount of replicase (figure 5a, dotted curve) from
which we computed the ratio of free positive strands to
replicase throughout generations. For the default replication
strategy, the computed ratio varies between 0 and 3.5
and stabilizes around 0.59 after some time (figure 5b, grey
curve). After synthesis of 12 polyproteins, each positive
strand is used immediately as a template for producing
two negative strands; hence the time lag and scale
difference between the two curves in figure 5b.

Our analytical model is in agreement with previous
findings (Regoes et al. 2005) that, in the absence of
encapsidation and mutation, the optimum balance of
translation and transcription consists of translating each
positive strand into 13 polyproteins, and then transcribing
each positive strand into two negative strands. In contrast,
our model predicts no upper limit to the number of
positive strands transcribed from each negative strand,
a finding which differs qualitatively from Regoes et al.
(2005), who do find an optimal value for \(n_P\) because they approximate the generation interval distribution by
its mean value (resulting in a growth rate expressed as a
ratio whose numerator increases logarithmically and
whose denominator increases linearly with \(n_P\)). However,
both results are quantitatively similar since their optimal
value of \(n_P\) (36.6) corresponds to more than 99 per cent
of the exact asymptotic growth rate. Including the
encapsidation process results in a decrease of approximately
5 per cent in the growth rate of virions. Interestingly,
when the capsid comprises numerous monomers
(e.g. \( n_0 \sim 2000 \) for the Potyviridae), the relationship \( R_N = r_N r_P (1 - r_N/n_0) \) implies that the impact of encapsidation becomes negligible on virion growth rate. In addition, the polyprotein strategy is one way, among others, to limit the rate of virus encapsidation, through producing one polyprotein for replication. The evolution of this combination of traits may have been driven by a selection pressure for high intracellular growth rates.

Replication leads to an exponentially increasing number of positive strands generated at each transcriptional generation. The addition of mutation to this process results in an exponential decrease in the proportion of genomes that are likely to be viable (figure 4a). When the mutation rate is high enough, the optimal set of parameters shifts towards a replication strategy with greater demographic contributions from the earlier generations (table 2), because the accumulated mutation load is lower. Surprisingly, our results suggest that the optimal replication strategy is broadly robust to the inclusion of mutation. Indeed, a mutation rate of \( 4.5 \times 10^{-4} \) mut/nt/rep reduces the growth rate of viable virions by 41–60% compared with no mutation (table 2), and the proportion of viable virions to 0.11 per cent at the time when 76 000 positive strands have been produced. However, maintaining an optimal growth rate of viable virions despite increasing mutation rates requires only a modest increase in the number of negative strands produced per positive strand and in the number of polyproteins synthesized prior to the onset of transcription (table 2). Indeed, most studies of Picornaviruses suggest that positive-sense viral RNA accumulates exponentially, at least in the early stages of cell infection (Novak & Kirkegaard 1991, 1994; Bolten et al. 1998; Li et al. 2009). Although we assumed that mutations are either lethal or neutral, the impact of slightly deleterious mutations might be considered to be cancelled by the impact of slightly beneficial mutations. More specific simulations of the effect of the fitness landscape on the observed mutation load have been conducted recently (Sardanyés et al. 2009).

Within a cell, the fittest virus might be considered as the one with the highest growth rate (as we assume here) or with the highest viral yield at a given time (which is equivalent, assuming a quick convergence towards the average growth rate). However, virus transmission between cells and at higher levels might depend on receptors that are present in limited numbers and can be blocked by non-viable virions; in such circumstances, the proportion of viable virions would be a key parameter. This would place within- and between-cell selection in conflict because the proportion and growth rate of viable virions are antagonistic for two reasons. First, there is a direct trade-off between fidelity and the polymerization rate of viral replicase (Furió et al. 2005, 2007). Second, our model has unveiled an evolutionary trade-off between the proportion and growth rate of viable virions: the strategy that maximizes the growth rate results in a minute proportion of viable viruses (figure 4a); increasing this proportion requires a move away from the optimum, towards a stamping-machine strategy, which reduces the number of transcriptional generations and thus the within-cell growth rate. In addition to this surprising role of mutation in the trade-off between the transmission and intracellular growth rate of viable virions, higher viral growth can lead to higher virulence, which is also assumed to trade-off with transmission in the classical models of virulence evolution (Coombs et al. 2003; Gilchrist et al. 2004; Alizon et al. 2009). Because of these three trade-offs, the resulting strategy might strike a compromise between the proportion and growth rate of viable virions (i.e. between quality and quantity of the progeny), or switch from a linear to an exponential replication strategy at some point in the cell-infection cycle. Aphid-transmitted plant viruses of the Potyvirus genus may be especially affected by this trade-off, because stylet-borne viruses initiate infections with so few virions (Moury et al. 2007) that the proportion of viable genomes might be crucial for transmission. Thus, we may expect such viruses to have evolved mechanisms that reduce the proportion of lethal genomes, while plant viruses transmitted after ingestion of large numbers of virions and circulation through their vector’s body may tolerate a higher proportion of lethal genomes.

The optimal combination of parameters in the absence of mutation can be intuitively explained in the following way. Translation and transcription are mutually exclusive events (Gamarnik & Andino 1998), and at least initially
(and here we have assumed this to apply throughout the replication cycle) translation must precede transcription, as replicase is required for transcription. Thus, delaying transcription will extend the virus generation time and reduce the growth rate, and therefore translation must be kept to the minimum required to provide the structural and non-structural proteins needed for transcription and encapsidation. A similar increase in \( R_0 \) can be obtained through a given increase in the number of genomes transcribed from either positive or negative strands. However, because the same amount of replicase (translated from the positive strands) has to be used first for transcription from these positive strands and then for transcription from the resulting negative strands, the limiting factors do not act in a symmetrical way on both strands: an increase in \( r_P \) results in a proportional increase in the total amount of available replicase, but any increase in \( r_N \) implies a higher demand for replicase, inducing either more translation or a shortage in replicase. As both of these outcomes extend the virus generation time, the best way for the virus to increase \( R_0 \) is through an increase in \( r_P \). In addition, negative strands are not required for anything other than positive strand synthesis, so this process can continue indefinitely (although the benefits of doing so become rapidly insignificant). Note that this explanation of the asymmetric replication does not involve encapsidation.

Our model predicts the time-averaged ratio of free positive to negative strands to be about 10, but suggests this ratio could be expected to fluctuate considerably in the early course of cell infection. In the model, these fluctuations arise from the discrete, deterministic and initially synchronized nature of the replication cycle. In reality, we would expect asynchronies induced by stochastic effects to damp out the oscillations much more quickly than shown in figure 5. However, it is important to note that contrary to what has been previously assumed (Regoes et al. 2005), the ratio of positive to negative strands is not expected to equate to the ratio \( r_P/r_N \) because the average excess in positive strands takes into account the time spent at different ratios over the whole replication cycle. For Poliovirus, a large excess (30–100) of positive strands has been measured in several studies (Andino et al. 1990; Novak & Kirkegaard 1991; Bolten et al. 1998; Paul 2002). Similar results have been obtained for other ssRNA(+)viruses, and there are indications of mechanisms that regulate the asymmetric production of positive and negative strands (Belsham 2005).

The ratio of free positive strands to replicase is predicted to be of the order of one-half, with considerable variability possible. However, the absence of any representation in the model of molecular degradation over time (through degradation either by the host proteins or by the effect of mutations) leads us to regard these ratio predictions with caution. More generally, we focused on the period of the replication dynamics during which viral genotypes actively compete, i.e. the exponential phase. Thus, we did not model the processes acting on the cell-infection cycle only at the beginning (when the virus gradually hijacks the cell machinery for its own replication) or at the end (close to the cell-carrying capacity, when nucleotides or amino acids or other components of the cell machinery can become limiting for virus replication).

Considerable uncertainty exists surrounding the rate of RNA virus mutation. Values reported can vary over three orders of magnitude (Holland et al. 1982; Parvin et al. 1986; Drake 1993; Duffy et al. 2008), and yet mutation rate is an important parameter fundamental to many questions about the evolutionary genetics of viral populations. One reason for the uncertainty in its measurement per genome replication is that while viral mutation frequency can be estimated with some precision, the number of generations of replication over which these mutants are generated is much harder to quantify. Indeed, even the definition of genome replication requires some care; it could be regarded as per transcription, or from positive strand to positive strand, or even from virion to virion. One possible future solution is to fit mutation frequency data directly to models of the sort developed here. Our increasing ability to manipulate and infect single cells, and experimentally modify mutation rates, combined with new high-throughput sequencing technologies, are likely to generate new opportunities to parametrize models of the population genetic dynamics within cells.

We thank Guillaume Martin for stimulating discussions, especially on the mutation model, and for helpful comments on the manuscript together with Sylvie Dialot, Serafin Gutiérrez, Gérard Labonne, David Pleydell and Virginie Ravigne.

Funding. This research has been funded by ANR-BBSRC SysBio (project EpiEvol).

REFERENCES


