The evolutionary biology of molecular parasites

S. NEE[1] *and* J. MAYNARD SMITH[2]

[1] *AFRC Unit of Ecology and Behaviour, Department of Zoology, South Parks Road, Oxford OX1 3PS*
[2] *Department of Biology, Sussex University, Brighton, Sussex BN1 9QG*

Key words: molecular evolution, transposons, coviruses, retroids.

INTRODUCTION

A parasite can be considered to be the device of a nucleic acid which allows it to exploit the gene products of other nucleic acids–the host organisms. In this view, all parasites are 'molecular parasites'. But it is interesting to restrict our attention to nucleic acids which do not encode organisms, as these live in a purely molecular world which lacks emergent features such as fangs and ovipositors. Viruses and transposons are molecular parasites in this sense. Most viral nucleic acids do code for some proteins, such as replicases and the protein shell in which they travel between their cellular oases. Some, however, do not even have a shell and code for nothing at all–these are the 'viroids' (Reisner & Gross, 1985), the smallest parasites in the world.

Transposons can be thought of as viruses that lack the ability to leave the cell in which they occur. They are able to make new copies of themselves and insert them elsewhere in the genome. This allows transposons to spread through the genomes of sexual species in spite of the fact that transposition may be harmful to the host (Charlesworth, 1987; Hickey & Rose, 1988). Harm may be a result of disruption of gene function at the site of insertion, for example, or a result of non-homologous recombination (Charlesworth, 1988). Selection at the level of the host for reduced transposition is a weak force compared to selection at the level of the transposon for increased rates of transposition (Charlesworth & Langley, 1986). (Because prokaryotes do not engage in sex very often, their transposons have a very different evolutionary biology which will not be discussed here. See Maynard Smith (1989) and Condit, Stewart & Levin (1988).)

Molecular parasites find themselves in the same situation, perhaps, as the molecular replicators in the primaeval soup; their gene products float freely in the cellular soup accessible to any other nucleic acid with the capacity to use them. In other words, these parasites *can use each other's gene products as well as those of the host*. As we will see, this simple fact has wide-ranging implications for their evolution and population biology.

Another important feature of molecular parasites is that many of them have very high mutation rates. There are error correction mechanisms associated with nuclear DNA replication which make it a very accurate process. There are no such mechanisms associated with RNA replication or reverse transcription, so RNA viruses (such as those that cause colds and flu), retroviruses (e.g. HIV) and transposons, such as copia and H.M.S. Beagle, that probably replicate by reverse transcription, are all subject to elevated mutation rates. The enormous variability of other transposons, like the LINES elements of mammals and the p elements of *Drosophila*, suggests that these elements may also be subject to high mutation rates, although little is known about their mechanism of transposition.

We will use the phrase 'mutation' in a loose sense. As well as 'point' mutation, which changes single nucleotides, it also encompasses the consequences of template jumping and switching by the replicase as it manufactures progeny nucleic acids. Deletions, duplications and recombinant progeny can be produced when the replicase jumps to a new position on the parent molecule or jumps to a new template altogether. These kinds of errors are commonly made by RNA replicases and are important in the generation of 'defective interfering' viruses which will be discussed below (Holland, 1985; Schlesinger, 1988). The kinds of replication errors that occur are determined not just by the replicase, but also by the cellular environment (e.g. Younger *et al.* 1981; O'Neill, Maryon & Carroll, 1982).

Just because a nucleic acid is DNA rather than RNA does not mean it has a mutation rate as low as that of organismic nuclear DNA. For example, mitochondrial DNA replication appears to be much more error-prone than nuclear DNA replication (Brown, 1983). The very small size of some DNA viruses, such as the parvoviruses of animals, and the intriguing geminiviruses of plants, suggests that their replication may be highly error-prone. This is because there is a direct relationship between mutation rate and genome size, as we will see. Parvoviruses produce 'defective interfering viruses' (see the Molecular Games section) in great numbers, which is direct evidence that their replication is highly error-prone (Berns, Muzyczka & Hauswirth, 1985).

The structure of the paper is as follows. The second section discusses the 'error threshold', that is, the limitation that is imposed on the size of the genome by the process of mutation. The formal derivation of this threshold is relegated to an

appendix, but the consequences for the sizes of viral genomes are discussed in the main text. We also analyse two processes–recombination and 'host jumping'–that may enable viruses to evolve larger genomes than would otherwise be the case.

The next section, on 'molecular games', discusses some consequences of the fact that molecular parasites have access to each others' gene products. The most obvious consequence is that viruses have viruses of their own: we describe some examples. We then turn to the fact that functional transposons often coexist in the genome with non-functional mutant derivatives, and discuss the possible relevance of this fact for transposon population biology. A related phenomenon is that of 'defective interfering viruses'–mutant viruses which, although defective when on their own, are superior to the wild-type when complemented. We describe two examples, the VSV system and the SV40 system. The latter leads on to a discussion of the 'coviruses' of plants, which we interpret as molecules that have lost some information essential for replication on their own, but which have a fitness advantage when their deficiences are complemented. Finally, we analyse a model, the 'retroid game', which may shed some light on the evolution of retrotransposons such as the copia element of *Drosophila*. The starting point of the game is a trade-off between producing RNA templates for replication, and hence leaving descendants, and producing messenger RNAs for protein production.

THE ERROR THRESHOLD

In an evolving population of nucleic acids, there will be some unique sequence, or set of very similar sequences, that have a higher fitness than any mutant: we will refer to this sequence as the 'wild-type'. Clearly, if the wild-type is to be maintained in the population, each wild-type individual must produce at least one wild-type offspring. This imposes a limit on the size of the genome for any given per-base mutation rate. This limit is the error threshold. If the replication error rate of a nucleic acid is above the threshold, the molecule cannot persist in the population. Very approximately, there must be a chance of 0·5 that an offspring be free of mutations or, equally approximately, that the number of bases in the genome should not be greater than the reciprocal of the per base mutation rate. A simple model is given in the Appendix, which should be consulted by those who wish to follow the other models in this chapter. For more detailed models and discussion, see Eigen & Schuster (1979) and Eigen & Biebricher (1988).

One consequence of the error threshold is that a molecular parasite whose replication accuracy falls below it may go extinct. (In the model given in the Appendix, even if W_A is larger than the degradation rate, d, when the molecules are rare, extinction will occur if QW_A and W_B are smaller than d.)

Another important consequence of the error threshold is the limit it sets on genome size (Eigen & Schuster, 1979). This is an obvious consequence of the fact that a replicase is more likely to make an error the longer the molecule it has to replicate (the more information, the smaller the Q parameter in the Appendix). The point mutation rate of RNA replication may be about 10^{-3} to 10^{-4} per base, compared to the rates of 10^{-9} to 10^{-11} that can be achieved for DNA (Domingo & Holland, 1988). As a result, DNA viruses can be far larger than RNA viruses. Vaccinia virus, the DNA virus used to eradicate smallpox, has well over 100 genes whereas a typical RNA virus has 4 or 5.

The sizes of molecular genomes are usually given as either their number of nucleotides or the molecular weight in Daltons (a hydrogen atom weighs 1 Dalton). One can switch between these measures by noting that a nucleotide pair of duplex DNA weighs about 660 Daltons, whereas a nucleotide of single-stranded RNA weighs about 345 Daltons (Coffin, 1984). Single-stranded RNA viruses typically weigh in at around 3–4×10^6 Daltons, whereas poxviruses (one of which is vaccinia) can exceed 200×10^6 Daltons. Coronaviruses, which cause cold-like symptoms in humans, have the largest monomolecular RNA genome; its weight is at most 8×10^6 Daltons. As noted in the Introduction section, many DNA viruses are very small. Hepatitis B virus has a genome weighing $1{\cdot}6 \times 10^6$ Daltons. It is now known that this virus actually replicates via reverse transcription from an RNA intermediate, so its replication is highly error-prone (Bosch, Kuhn & Schaller, 1988). (Table 1 presents information about the important viruses mentioned in this chapter. A convenient summary of basic facts about all viruses has been provided by Matthews (1979).)

The error threshold limitation on genome size has led to the origin of life catch-22 for nucleic acid replicators: no accurate replication without proteins and no proteins without accurate replication (Eigen & Schuster, 1979). We will now discuss the possibility that a variant of the paradox might be involved in the explanation of the striking absence of error correction mechanisms in RNA replication systems. Nucleic acid replication unassisted by any enzymes at all has an error rate of around 10^{-2} per nucleotide, i.e. one nucleotide in a hundred is mismatched. A replicase (or 'polymerase', as they are called for DNA) reduces the error rate to around 10^{-5}, perhaps by increasing the mean free energy difference between matched and mismatched nucleotides (Weissbach, 1977). Polymerases differ in their accuracy (Weissbach, 1977; Kornberg, 1980). They also differ in whether or not they have exonuclease activity. Exonuclease activity is proof-reading: an incorrect nucleotide is removed before replication

Table 1. Viruses discussed in text

(Most of the following information is from Fields (1985) and Matthews (1979). Unless otherwise indicated, the genome consists of a single molecule of nucleic acid which may be single-stranded, ss, or double-stranded, ds. Single-stranded RNA genomes may be either positive-sense, +, or negative-sense, −. Positive-sense RNA molecules are treated as messenger RNAs by the cell, whereas negative-sense molecules must first be copied into positive-sense molecules by a virus replicase. The genome weight is in Daltons.)

Virus	Comments	Genome	Genome weight
Hepatitis B virus	Replicates via reverse transcription	ds DNA	1.6×10^6
Parvoviruses	Widespread in warm-blooded animals	ss DNA	$1.5–1.8 \times 10^6$
Vesicular stomatitis virus	Causes oral lesions in cattle	ss RNA, −	3.9×10^6
Simian virus 40 (SV40)	A papovavirus	ds DNA	3.5×10^6
Vaccinia virus	Used to eradicate smallpox	ds DNA	123×10^6
Coronavirus	Causes cold-like symptoms in humans	ss RNA, +	$5–8 \times 10^6$
Poliovirus	A picornavirus	ss RNA, +	2.5×10^6
Retroviruses	e.g. HIV 1	ss RNA, + two identical copies per virion	$< 3.4 \times 10^6$ (each copy)
Influenza virus	An orthomyxovirus	ss RNA, − 8 components	5×10^6 (total)
Reovirus	Wide host range	ds RNA 10–12 components	$12–20 \times 10^6$ (total)
Cystovirus	Infects bacteria	ds RNA 3 components	10.4×10^6 (total)
Adenovirus	Causes respiratory ailments in soldiers	ds DNA	24×10^6
Adeno-associated virus	A parvovirus parasitic on adenovirus	ss DNA	$< 1.8 \times 10^6$

proceeds. Proof-reading is present in the DNA polymerases of prokaryotes, yeast and vaccinia virus for example (Kornberg, 1980), but is absent from RNA replicases and has not been widely observed in eukaryotes. (Eukaryotic nuclear DNA replication achieves its extremely high accuracy from an additional mechanism of post-replication mismatch repair (Modrich, 1987).)

The origin of life catch-22 arises because a molecule with a replication error rate of 10^{-2} cannot encode enough information for a polymerase due to the error threshold. It must have a polymerase if its accuracy is to be high enough to allow it to encode a polymerase. Possible resolutions of this problem are the hypercycle (Eigen & Schuster, 1979) and the stochastic corrector (Szathmary, 1989), involving intermolecular cooperation and group selection, respectively. But, having solved this problem, there may be a new one: perhaps the molecule faces a new error threshold that prevents it from evolving a very accurate polymerase or one with exonuclease activity. We entertain this speculation because polymerases are very large, complex proteins composed of multiple subunits. One could explore its plausibility by investigating the information requirements of enhanced accuracy and exonuclease activity.

A troublesome fact for this speculation is that the RNA polymerases which manufacture messenger RNAs from DNA also appear to lack proof-reading. But the DNA genomes of organisms clearly have no error threshold limitation preventing them from equipping their RNA polymerases with exonuclease activity. But before we conclude that there is just something about RNA chemistry which rules out proof-reading, we must bear in mind that the information for a particular protein is only a fraction of the whole genome and the messengers are continually being manufactured and degraded. So it may be that there is simply no need for proof-reading.

Similarly, Eigen & Biebricher (1988) have suggested that small viruses, like RNA viruses, simply do not require an extremely accurate replicase. This raises the question which can be put as follows: do small, low-accuracy replicators have a similar *genomic* mutation rate to big, high-accuracy replicators? We suspect the answer is no, because there are three lines of evidence suggesting that viruses with inaccurate replicases are paying a price.

Table 2. The error threshold with recombination

(There are 5 loci, and back mutation is ignored. In the multiplicative case, fitness is $(1-s)^k$, where $s = 0{\cdot}05$ and k is the number of mutant loci. In the synergistic case, fitness is $(1-s)^{k^2}$: in the submultiplicative case, fitness is 1 for the wild-type and $1-s$ if $k \geqslant 1$. In the cases with recombination, pairing is random, and each offspring has one randomly situated cross-over.)

Type of mutation interaction	Values of the per-locus mutation rate above which the frequency of the wild-type would go to zero	
	No recombination	With recombination
Multiplicative	0·0256	0·0256
Synergistic	0·048	0·29
Submultiplicative	0·0098	0·0016

(1) Low-accuracy viruses are non-infectious to a great extent. Almost all the virions of DNA bacteriophages are infectious (Luria *et al.* 1978) whereas just one half of the RNA bacteriophage virions are unable to inject their RNA (Lewin, 1977). The situation is even more extreme among animal viruses 1 in 4 vaccinia (DNA) virus virions may be infectious (Luria *et al.* 1978) whereas only 1 in 100 hepatitis B virions is infectious (Coffin, 1988) and even fewer poliovirus virions (Luria *et al.* 1978). (Luria *et al.* (1978) reported that 1 in 2·6 virions of mouse encephalitis virus, an RNA virus, is infectious, but they misread the paper cited in support of this: (i) the virus is called mouse encephalo*myelitis* virus; (ii) the ratio 1:2·6 refers to plaque-forming units: mouse infective particles, which is not the same thing at all.) (2) Low-accuracy viruses are often very economical in the way they encode their genetic information; the same stretch of molecule may code for pieces of different proteins in different reading frames. All the retroviruses have overlapping reading frames to a limited extent and hepatitis B virus exhibits this phenomenon to a great extent: about 50% of the genome is read in more than one frame and, within a frame, multiple start codons are used to generate multiple, related proteins (Ganem & Varmus, 1987). This kind of genetic organization has at least one obvious drawback: it constrains the adaptation of proteins that share a coding domain. Overlapping reading frames are also found in the following RNA viruses: Bunyaviruses, influenza virus, Sendai virus and respiratory syncytial virus perhaps (see the relevant chapters in Fields, 1985). (3) Small replicators produce 'defective interfering viruses' in great abundance (see the 'Molecular Games' section).

We will now discuss two ways of getting around the error threshold: recombination and host jumping.

Recombination

At first sight, it would seem that recombination should enable a population to sustain a higher load of mutations, because it makes possible the construction of one functional molecule out of two defective ones. Things are not so simple, however, because recombination can also break up a functional molecule. It turns out that the effect of recombination depends critically on how mutations combine in determining fitness, w, which, in turn, determines the shape of the mutation distribution. The simplest assumption is the multiplicative one: i.e. $w = (1-s)^k$, where k is the number of mutations and s is the selection coefficient against a mutation. If this is the case, then recombination has no effect on the error threshold. If, however, mutations interact 'synergistically', so that each additional mutation leads to a greater proportional decrease in fitness, then recombination can increase the permissible mutation rate or, equivalently, the permissible genome size for a given mutation rate (Kondrashov, 1988). In contrast, if mutations interact 'submultiplicatively', i.e. additional mutations produce little further reduction in fitness, recombination *reduces* the permissible mutation rate. Table 2, obtained by simulation, shows that the effects can be large.

There is evidence that recombination does allow viruses to be larger. Some RNA viruses, such as influenza virus, are multicomponent, i.e. their genome consists of more than one RNA molecule (8 in the case of influenza virus). Essentially, these viruses have more than one 'chromosome'. If a cell is infected by two such viruses, a recombinant progeny virion may be produced containing segments from each parent. It is known that reassortment does, indeed, occur *in vivo* (Gombold & Ramig, 1986; Stott *et al.* 1987).

Pressing & Reanney (1984) drew attention to the fact that multicomponent virus genomes are larger than monomolecular RNA genomes. For example, the reovirus genome, with 8–12 segments, has a total molecular weight of $12–20 \times 10^6$ Daltons, which is unusually large even if we allow for the fact that the

RNA is double-stranded in this virus. If this large size is due to an ability to transcend the error threshold, as Pressing & Reanney (1984) suggested, then recombination may be the source of the ability. There are other possibilities, however. Pressing & Reanney (1984) argued that miscopied RNA molecules may be less effectively encapsidated, i.e. there is a form of 'proof-reading' due to the mechanics of the assembly of new virus particles. These possibilities are not mutually exclusive, of course. We believe that recombination is probably an important factor because of the fact that coronavirus, the largest monomolecular RNA virus, exhibits an extraordinarily large incidence of recombination, almost as high as multicomponent viruses (Lai, 1988).

But we are unable to exclude the possibility that the large size of multicomponent virus genomes (and the small size of single component genomes, for that matter) has little to do with the error threshold, but is mainly due to the fact that multicomponent viruses are better able to deal with the monocistronic messenger problem (Reanney, 1982). The protein production machinery of eukaryotes is only equipped to translate monocistronic messenger RNAs, so an RNA virus has to have some device to produce its several proteins. There is a variety of ways of doing this (Roizmann, 1985). Postive sense RNA genomes (those which can function as messengers) may be translated into polyproteins which are subsequently cleaved. This occurs in poliovirus, for example. Viruses with negative sense RNA genomes may use their own RNA polymerase to transcribe subgenomic messengers. Vesicular stomatitis virus uses this tactic. However, just because there are ways of dealing with this problem does not mean it is not still a problem constraining the size of a monomolecular genome. This problem is less severe, of course, for a multicomponent virus which has its genome divided up on different RNA molecules.

Pressing & Reanney (1984) presented a very different argument against the view that the large size of multicomponent genomes is simply due to the monocistronic messenger problem. They observed that cystovirus, a multicomponent double-stranded RNA virus of bacteria, also has a large genome compared to single component RNA bacteriophage (total weight $10{\cdot}4 \times 10^6$ Daltons compared to $1{\cdot}5 \times 10^6$ for the single-stranded monomolecular RNA phages), but that bacteria are quite capable of processing polycistronic messengers. This seems to be a powerful argument.

Before leaving the subject of multicomponent viruses we must note a striking fact which we will not attempt to explain here: all double-stranded RNA viruses are multicomponent viruses.

Retroviruses, which are single-component viruses, also engage in recombination to a great extent (Linial & Blair, 1984). This may be facilitated by their unusual genetic structure. The retrovirus genome is contained on a single, linear RNA molecule, but each virion contains *two* such molecules which are joined at one end (Coffin, 1984). It seems that the reverse transcriptase only produces one copy of the genome, the 'provirus' (Coffin, 1988; Panganiban & Fiore, 1988), which is integrated into the host DNA. It is plausible that recombination is achieved by the reverse transcriptase switching templates during the manufacture of the provirus.

Recombination is not observed in single-component RNA viruses, apart from those we have mentioned, retroviruses and coronavirus, nor in picornaviruses, such as poliovirus (King, 1988). Why not? An obvious explanation is that recombination is not favoured by selection, perhaps because the shape of selection is submultiplicative. But even if recombination would be advantageous, single-component RNA viruses may simply not be able to do it due to their inability to code for the large number of enzymes necessary to carry out homologous recombination reliably. It is thought (King, 1988) that RNA recombination is effected by a copy-choice mechanism, whereby the replicase switches templates as a progeny molecule is being constructed. Such a mechanism will produce deletions or insertion mutations if the replicase does not pick up where it left off, which must be a daunting task for such simple molecular systems. So any advantage there might be in elevating the recombination rate must be large enough to compensate for the disadvantage of producing massively mutated offspring at an elevated rate as well. However, poliovirus is reported to have evolved a mechanism of reliable homologous recombination (King, 1988) which suggests this problem is not insoluble.

Host jumping

Another way around the error threshold may be host jumping. Suppose that, typically, a new host is infected by one or a few viruses (or by the recent descendants of one or a few viruses). This can result in the maintenance of the wild-type virus, even if, within a host, the mutation rate is above the error threshold. This will occur if the wild-type has an advantage in transmission to *new* hosts, so it can escape the accumulating mutants *within* a host before being overwhelmed. We will now present a model of this process.

Suppose that hosts are initially infected by a single virus which may be W (wild-type) or M (mutant). We are interested in the number of new hosts infected by viruses arising from this infection. In the case of a wild-type virus, this number is R; this reproductive rate is the 'fitness' of a single infection. If the original infection was by a mutant virus, the 'fitness' is $R^o(1-s)$. Thus s measures the selective

disadvantage, over the whole period of infection, of a population of mutant viruses.

If the infection was by an M virus, all transmitted viruses are M (that is, we ignore back mutation). If the infection was by a W virus, the transmitted viruses are in the proportions $kM:(1-k)\,W$. Thus k measures the rate at which mutants rise and spread through the virus population within a host.

Let the initial frequencies of W and M infections be $p:(1-p)$. The W infections give rise to $pR(1-k)$ new W infections, and pRk new M infections, and the M infections give rise to $(1-p)\,R(1-s)$ new M infections. The total of new infections is $pR+(1-p)$ $R(1-s)$ and of W infections is $pR(1-k)$.

To keep things simple, we will suppose that all new hosts are infected at the same time; p is the frequency of W infections in the present hosts and p' is the frequency of W infections in the next batch of hosts. We have

$$p' = \frac{p(1-k)}{1-s(1-p)}. \quad (1)$$

If an equilibrium with $p \neq 0$ exists, then $p' = p$, or

$$p = 1-\frac{k}{s}. \quad (2)$$

Thus, the maintenance of wild-type viruses requires that $k < s$. That is, the spread of mutants within a host must be less than the between-host advantage of the complete virus.

Now consider the case in which a host is simultaneously infected by n viruses (if there is a substantial time-interval between infections, only the first infection is relevant, and the simple model applies). As before, the proportion of M viruses transmitted, following an all-W infection, is k. For simplicity, we assume that a mixed infection is equivalent to an M infection: this extreme assumption makes it more difficult for host-jumping to maintain W. If, when infection occurs, the n virus particles are drawn randomly from the previous host, the probability of an all-W infection is $(1-k)^n$. Thus the only modification called for is to replace $(1-k)$ by $(1-k)^n$ in equation (1). The condition for the maintenance of wild-type viruses is then

$$1-(1-k)^n < s.$$

As n becomes large, $(1-k)^n \to 0$, and the condition is hard to satisfy.

We conclude that host-jumping can maintain wild-type viruses, even if the mutation rate is above the error threshold within a host, provided n is not too large, and that k is not too large relative to s. Note also that if a virus has been able to evolve a larger genome for this reason, it would not then be sensible to evolve direct transmission to the next host generation via the host egg; this may be one reason why the direct transmission of viruses is unusual.

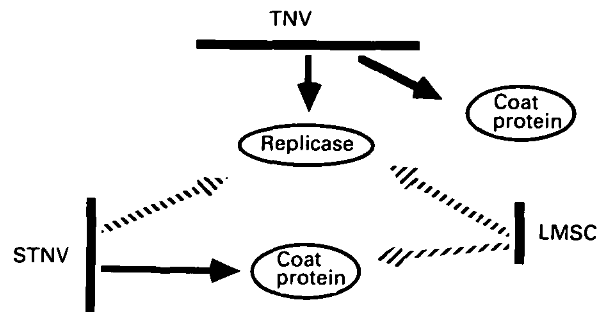

Fig. 1. Relationships among tobacco necrosis virus (TNV), satellite tobacco necrosis virus (STNV) and low molecular weight satellite component (LMSC). The length of the lines representing the viruses indicates the relative amounts of functional information. A solid arrow means that the virus both encodes and uses the indicated protein. A broken arrow means that the virus just uses the protein.

More generally, we wonder if the very short periods of host infection that characterize highly error-prone viruses, such as influenza virus, might be partly a host-jumping virus strategy.

MOLECULAR GAMES

We emphasized earlier that molecular parasites have access to each others' gene products. We now discuss some consequences of this fact.

Viruses of viruses

Viruses often have viruses of their own. For example, tobacco necrosis virus (TNV) is an RNA plant virus which is parasitized by a virus called satellite tobacco necrosis virus (STNV) (Franki, 1985). STNV codes for its own coat protein, but depends on TNV for the replication of its RNA. The relationship is one of exploitation since the production of TNV in plants coinfected with STNV can be so depressed that TNV becomes serologically undetectable. Like the big fleas with little fleas, STNV has its own parasite, low molecular weight satellite component, LMSC (Frank, 1985). LMSC uses the replication machinery of TNV and the coat protein of STNV: it is, in fact, a virus of a virus of a virus. These relationships are illustrated in Fig. 1.

Transposons and complementation

For many transposons, the number of elements per host genome is large, and recombination is a very effective mixing process, so wild-type and mutant elements coexist and the mutants have a high probability of having their deficiencies complemented.

Consider mutants which have non-functional coding regions (perhaps they are deleted) but retain

the recognition sites necessary to interact with replicases, transposases etc. Such molecules are incapable of replicating on their own but may replicate just as effectively as the wild-type if provided with the necessary proteins. It is at once obvious that such mutants may accumulate if complementation probability is high. Not only are they continually being generated by mutation, but they are also being replicated with the assistance of wild-type molecules.

Not surprisingly, in view of the above, the members of a transposon family are typically highly heterogeneous in size and sequence and many are known to be both defective and transposable in the presence of functional elements (for reviews of transposons, see Spradling & Rubin (1981) and Shapiro (1983)).

It is reasonable to suppose that the total number of copies of a particular transposon cannot increase to infinity. One way the numbers may be regulated is by selection against hosts with a large number of elements due to the non-homologous recombination they induce (Charlesworth, 1988). In this case, the accumulation of completely defective mutants raises the distinct possibility of extinction of the transposon. A simple model is the following. Let F_A and F_B denote the 'fitnesses' of wild-type and mutant, where fitness here means the number of *correct* copies made of an element. R is the probability that a mutant is complemented and W_B is the number of copies made of it upon complementation. Using the symbolism of the Appendix,

$$F_A = QW_A$$
$$F_B = RW_B. \quad (4)$$

The mutants will certainly increase in frequency if $F_B > F_A$, i.e.

$$R > \frac{QW_A}{W_B}. \quad (5)$$

As noted above, this may be an easy condition for transposons to satisfy. When the wild-type element has become very rare, a stochastic treatment is necessary to determine its ultimate fate. Such an analysis was carried out by Kaplan, Darden & Langley (1985) who concluded that the ultimate fate is extinction. They noted that this result is consistent with the narrow taxonomic distribution of many transposons, which argues against an ancient evolutionary history. (Speculation about the evolutionary *history* of transposons and viruses is a fascinating growth area which is beyond our scope. See Doolittle *et al.* (1989) and Zimmern (1988).)

Host countermeasures against transposons

We may ask the question, 'what kinds of mutation in the host genome are likely to help against transposons?' A direct route to the evolution of novel weapons against transposons would appear to be via mutations in a transposon itself which, for example, causes it to produce a transposase which simply binds to the transposons, or excises them, without transposition. From the point of view of whose interests are being served, such elements can be thought of as having defected to become host genes, i.e. their spread through the host population is now due to selection at the level of the host rather than at the lower level of selfish DNA. This is the molecular version of the principle 'set a thief to catch a thief'.

There are no solid examples of this at the moment. However, it is suspected that a product of defective p elements may be what provides protection against the unbridled transposition which causes P–M hybrid dysgenesis in *Drosophila* (Simmons *et al.* 1987). It is also suspected that a mutant mariner element is responsible for the enhanced excision of that element from the genomes of some lines of *Drosophila* (Bryan, Jacobson & Hartl, 1987), although this may, in fact, be a byproduct of an enhanced rate of transposition.

So far, we have assumed that the fitness of the mutant is less than or equal to the fitness of the wild-type (i.e. $W_B \leqslant W_A$). In the discussion of defective interfering viruses (below) we will see that, in fact, a molecule may trade self-sufficiency for a selective *advantage* when its deficiencies are complemented (i.e. it may be that $W_B > W_A$). This has many consequences. One of these is a counterintuitive explanation of the anomalous antiquity of the copia elements of *Drosophila*, as inferred from their relatively widespread taxonomic distribution (Kaplan *et al.* 1985). This will be discussed below under the heading 'the retroid game'.

Defective interfering viruses

Satellite viruses, such as STNV and adeno-associated virus (a parasite of human adenovirus) appear to be unrelated to the viruses they parasitize. Defective interfering viruses (DI viruses), which we will now discuss, are mutant progeny of their viral helper. DI viruses are defective because they are lacking some essential genetic information (they are typically deletion mutants). They are 'interfering' because they interfere with the replication of the wild-type virus, by outcompeting it for the replicase, for example. When viruses are grown under conditions of high multiplicity passage in the laboratory, DI viruses rise to high frequencies. In high multiplicity culture conditions, DI viruses are complemented by wild-type virus, which provides the proteins they lack. The amplification of DI viruses has many consequences for virus titre, infectiousness and virulence, typically lowering all three. DI viruses are probably important in the establishment of persistent infections and other

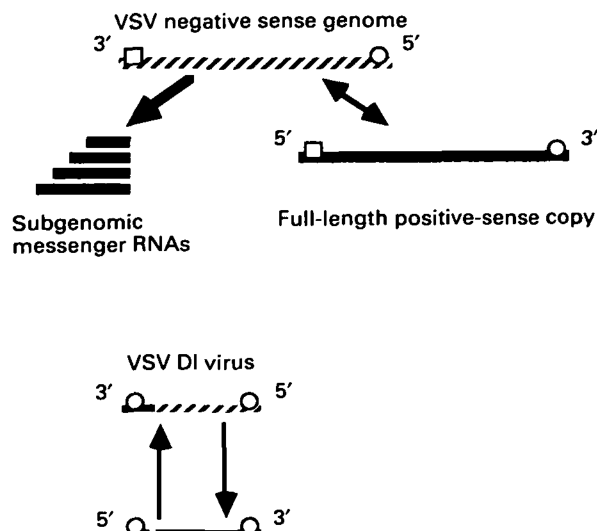

Fig. 2. VSV and VSV-DI virus genome structure. The negative-sense vesicular stomatitis virus genome is both transcribed into positive-sense messenger RNAs, which are smaller than the full genome, and replicated into a full-length positive-sense copy, which serves as the template for the production of progeny, negative-sense genomes. Both replication and transcription proceed in the 3′–5′ direction. The square is an initiation site for both transcription and replication. The circle is an initiation site for replication only. The defective interfering VSV has transposed the 3′ end of a positive sense molecule to its 3′ end, so the VSV-DI virus is only replicated. (Adapted from Holland, 1985.)

features of virus disease (e.g. Huang, 1988). They have been observed in all virus systems where they have been sought, with the exception of vaccinia virus. For interesting guides to the literature on DI viruses, see Holland *et al.* (1982) and De Polo, Giachetti & Holland (1987). For a review of the molecular processes which generate DI viruses, see Holland (1985) and Schlesinger (1988).

DI virus genomes are only defective in that they are lacking some important genetic information. In fact, they enjoy a selective *advantage* over wild-type virus in high multiplicity conditions where they are complemented. The reasons for the selective advantage are varied (Re & Kingsbury, 1988; Holland, 1985). Simply being shorter may itself be an advantage if this results in faster replication (Spiegelman, Mills & Kramer, 1975). But there are limits to how short a DI molecule can become, perhaps determined by the requirements of encapsidation. For example, Sendai virus is about 15000 nucleotides long and 1600 nucleotides appears to be the smallest its DI molecules can get before sacrificing their advantage (Re & Kingsbury, 1988). The DI molecule of tomato bushy stunt virus is only 396 nucleotides: it can constitute as much as 60% by weight of the virus nucleic acid in an infected plant, but constitutes less than 4% of the encapsidated RNA (Hillman, Carrington & Morris, 1987).

Two examples will show the more subtle ways in which there is a trade-off between being defective and having a selective advantage when other molecules can compensate for the deficiencies. **(1) The VSV system.** Vesicular stomatitis virus, which causes oral lesions in cattle, is a single-stranded negative-sense RNA virus (Emerson, 1985). Because its genome is negative sense, it must be transcribed into positive-sense messenger RNA. (If a virus has a positive-sense RNA genome, the genome can function as messenger RNA.) The genome is replicated by being copied into full-length positive-sense copies which then serve as templates for the production of negative-sense progeny genomes. The production of messengers and the production of positive-sense progeny intermediates are not the same thing because the messengers are shorter than the whole genome (Fig. 2). As illustrated in Fig. 2, the 3′ end of the negative-sense molecule has both a transcriptase and replicase recognition site, whereas the 3′ end of the positive-sense molecule has only a replicase recognition site (recall that both transcription and replication of nucleic acids proceed in a 3′–5′ direction along the template). As illustrated in Fig. 2, in addition to being shorter, one kind of VSV-DI has done away with the transcriptase recognition site by transposing a 3′ end from a positive-sense molecule (Holland, 1985). Effectively, the DI molecule does not waste any time being transcribed, but is simply replicated. A quantitative manifestation of this is that negative- and positive-sense DI molecules are synthesized in a 1:1 ratio compared to a 4:1 ratio for wild-type VSV (Emerson, 1985). **(2) The SV40 system.** Simian virus 40 is a small double-stranded DNA virus (Shah, 1985). SV40 DI molecules are the same size as SV40, but have sacrificed some genetic information to make room for a multiplication of the replication origin sequences. The rate of SV40 DI replication relative to SV40 replication is an exponentially increasing function of the number of origin sequences (Lee & Nathans, 1979).

Although there is no evidence of it in nature, perpetual host–parasite coevolution is a popular phenomenon amongst theorists as it may be a reason why most organisms reproduce sexually (references in this volume and Nee, 1989*b*). An interesting example of apparently perpetual host–parasite coevolution comes from the VSV/VSV-DI system. De Polo *et al.* (1987) maintained VSV under conditions of high multiplicity serial passage for over 2 years (more than 500 passages). They observed that the dominant VSV molecule at any one passage was highly resistant to interference by the DI molecule that was dominant a few passages earlier but highly susceptible to the dominant DI molecule a few passages later. Sequence analysis showed that the coevolution appears to be occurring in the nucleocapsid protein and the encapsidation sequences at the 5′ end of the molecules. This

suggests that VSV escapes parasitism by evolving a new 'lock and key' for incorporation into nucleocapsids and the DI VSV subsequently evolves the right key.

Another interesting phenomenon has been observed in the SV40/SV40-DI system (O'Neill *et al.* 1982). In a particular cell line, two kinds of DI SV40 that are able to complement each other eliminated wild-type SV40. One kind of DI molecule sacrificed the 'early' genes and the other sacrificed the 'late' genes. Together, they propagate efficiently and rapidly. As O'Neill *et al.* (1982) noted, the phenomenon they observed is reminiscent of the 'coviruses' of plants, which we will now discuss. (Two strange features about this system must simply be noted. (1) SV40 DI molecules arise during *low* multiplicity serial passage. (2) The complementing DI molecules exhibit anomalous infectivity kinetics.)

Coviruses

Coviruses are so called because no single virion contains all the necessary information for successful infection. For example, there are two kinds of tobacco rattle covirus virion–a long particle and a short one. The RNA genome carried in the long particle codes for the replicase, while the coat protein gene is carried by the short particle. Successful infection in nature requires that cells be co-infected by both kinds of virion (Bruening, 1977). Coviruses are very common; 7 of the 19 groups of plant viruses are coviruses. (We have not counted the enigmatic geminiviruses as coviruses.)

If we suppose that coviruses have evolved from viruses which contained all necessary genetic information, there are two factors that are relevant for understanding their evolution. **Factor 1**. Molecules which are missing some information have a higher replication fidelity when they are copied because they have less information to be accurately copied. **Factor 2**. Molecules which are missing some information may have exchanged it for a fitness advantage when their deficiencies are complemented, i.e. they may be DI molecules. Maynard Smith (1989) modelled the evolution of coviruses considering factor 1 alone. This analysis shows that if the complementation probability is sufficiently high, then a pair of mutually complementing incomplete molecules can eliminate a complete molecule from the virus population. If we define fitness as the number of *correct* copies made of a molecule, then the incomplete molecules have a higher fitness, when complemented, as a result of their higher replication fidelity.

Nee (1987) was particularly interested in factor 2. Not surprisingly, mutually complementing DI viruses can replace a complete virus if complementation probability is sufficiently high. Less obviously, the very existence of DI viruses which code for nothing at all implies that the molecules in a covirus *must* be DI viruses which retain some functional information, at least in the symmetrical model studied. Let C, I, J denote complete, incomplete and junk molecules. Junk molecules code for nothing. We consider the fate of two kinds of incomplete molecules that can complement each other (Fig. 3). K_C, K_I and K_J are the number of copies (not necessarily correct) made of a molecule if it is complemented. The complementing incomplete molecules are assumed to have all the same parameters. Nee derived a necessary condition for the evolution of the covirus, i.e. the elimination of the complete molecules (he mistakenly presented this as a strict inequality):

$$K_C K_J \leqslant K_I^2. \tag{6}$$

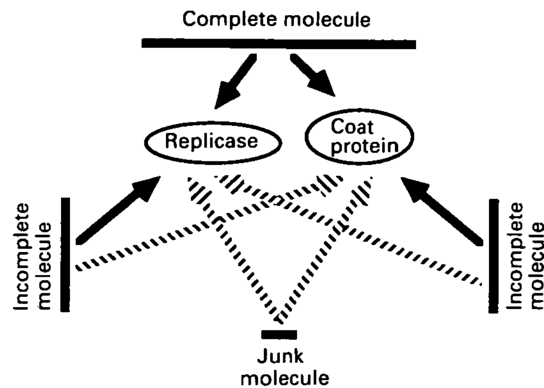

Fig. 3. Relationships among complete, incomplete and junk molecules. For simplicity, we suppose that there are only two relevant proteins, a coat protein and a replicase. See the legend to Fig. 1 for the symbolism.

In this model, we are free to restrict our attention to DI junk, so we know that $K_J > K_C$. But then condition (6) tells us that for a covirus to evolve it must be the case that $K_I > K_C$, i.e. the incomplete molecules are also DI molecules. It would be interesting to know what happens in more complex models. For example, can a covirus evolve if only one of the complementing molecules is a DI molecule? If the answer is no, then coviruses can only evolve from viruses that produce diverse, mutually complementing DI viruses.

In this region of parameter space, $K_I > K_C$, coviruses can evolve no matter how high the replication fidelity of the complete molecules is (Nee, 1987). This raises the question of the relative importance of factors 1 and 2 in the evolution of coviruses. At first sight, the fact that all coviruses are RNA viruses suggests that factor 1 is quite important. But another generalization is that all coviruses are *plant* viruses. It may be that it is the plant connexion which is important, in which case the fact that all coviruses are RNA viruses may simply be due to the fact that most plant viruses are RNA viruses. It would be interesting to compare

animal and plant virus transmission multiplicities both between hosts and between cells within a host. We expect that virus infections of plants enjoy higher multiplicities at one or both of these stages and that this is why all coviruses are plant viruses.

Even if we decided that factor 1 was irrelevant and factor 2 all important, the low replication fidelity of the virus genomes may still be important in the evolution of the coviruses for a different reason. For a covirus to evolve, the incomplete DI viruses must come into existence. If we visualize DI molecules scattered around in an ill-defined nucleic acid space, it may be that the cloud surrounding the wild-type virus will be too compact to encompass the DI viruses if replication fidelity is very high. Recall that the only virus which has not been observed to spawn DI viruses is vaccinia virus; the large size and polymerase exonuclease activity of this virus suggest that it has a high replication fidelity.

With the notable exception of Reanney (1982), coviruses and multicomponent viruses (recall the recombination subsection) have traditionally been considered simply to represent two different ways a virus can have a divided genome, and the covirus phenomenon has been explained in terms of the advantages of recombination that accrue to a virus with a divided genome. We find this approach conceptually problematic (Nee, 1989*a*) and prefer to consider the evolution of coviruses and multicomponent viruses separately.

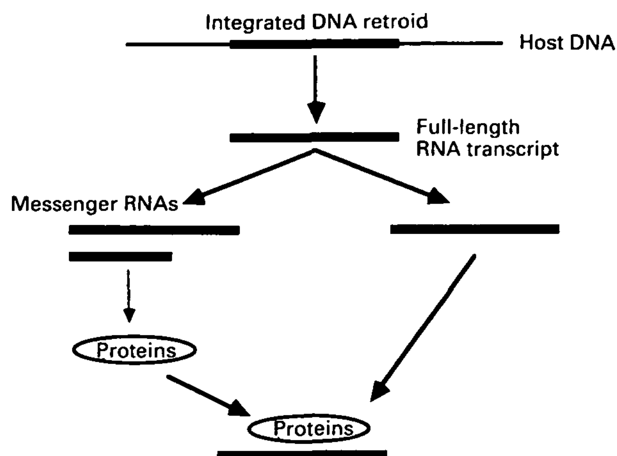

Fig. 4. Idealized retroid (retroviruses and retrotransposons) biology. The full-length primary transcript is either processed into messenger RNAs for protein production, by splicing or assignment to the appropriate cellular 'pool', or remains as a full-length template for the production of progeny retroids. In retroviruses, the proteins and templates combine to form progeny virions. In retrotransposons, the templates are reverse transcribed into DNA and integrated elsewhere in the genome.

The retroid game

With particular reference to retrovirus biology, we will now consider a way molecules may trade some self-sufficiency for an advantage when they are complemented. As our considerations encompass retrotransposons as well, we will use the generic term 'retroid' (Fuetter & Hohn, 1987). Retrotransposons, such as copia of *Drosophila*, are transposons which have a genomic structure that makes it obvious they are closely related to retroviruses. Little is known of their biology, but there are enough parallels with that of retroviruses that we will assume their biology is the same in its essentials (Boeke, 1988).

Consider the idealised retroid biology illustrated in Fig. 4. Full-length primary transcripts have one of two fates. They are either processed into messengers for protein production or serve as templates for the reverse transcription production of progeny. Depending on the ultimate product, processing into messengers may involve either splicing or simply the assignment of full length transcripts to the appropriate cellular 'pool' (Varmus & Swanstrom, 1984; Levin & Rosenak, 1976).

A retroid has to make an evolutionary decision as to the proportion of its primary transcripts it should shunt into protein production. (It is known for retroviruses that the proportion spliced is at least partly under the control of the retrovirus; Katz, Kotler & Skalka, 1988.) As evidenced by the complementation of defective retroviruses (Katz *et al.* 1988; Overbaugh *et al.* 1988; Lowy, 1985), retroids have access to each others' gene products. So it is clear that this decision will be affected by the protein production of the other retroids in the cell. For example, if a retroid commonly finds itself in cells with a large number of other retroids producing large amounts of proteins, then it may increase its replication rate by preserving a larger proportion of its transcripts as templates for progeny production. The decision a retroid has to make is the degree to which it should exploit the proteins made by other elements versus sacrificing some of its own templates to make proteins.

We will construct a simple model of the evolutionary decision-making process. We hope that this model exhibits qualitative features which are independent of the precise details of the molecular biology. When such details become available, we expect to see more realistic models making quantitative predictions.

Suppose that there are always $N+1$ retroids in a cell playing strategy p^* (i.e. they put a fraction p^* of their templates into protein production). Consider a mutant element playing p instead. We seek the ESS p^* such that the mutant will have a lower replication rate than the other elements if $p \neq p^*$. We will suppose that the rate of replication of a retroid, r, is proportional to the product of the concentration of

its templates and the concentration of proteins in the cell:

$$r \propto (1-p)(p^*N+p). \quad (7)$$

We also assume that selection at the level of the retroid maximizes the replication rate, r. This is reasonable for transposons unless the deleterious consequences of transposition are very large (Charlesworth & Langley, 1986). In any case, it is interesting to see what the outcome is for replication rates if retroids are parasitizing *each other* as part of a strategy for maximizing their replication rate.

To find the ESS p^*, we differentiate r with respect to p, set the derivative equal to zero and set $p = p^*$. This gives

$$p^* = \frac{1}{N+2}. \quad (8)$$

This says that the fraction of templates a retroid turns over to protein production declines as the typical number of elements in the cell increases.

There are two important effects of this mutual parasitism. **(1)** If retroids ignored the presence of other retroids and simply played the p that would be optimal if they were always alone, for example, the replication rate of each would be increased by a factor proportional to N. But by playing the ESS, the rate of replication per retroid is proportional to $(N+1)^2/(N+2)^2$. This quickly asymptotes at 1, which means that the retroids do not enjoy the significant replication enhancement they otherwise would. This outcome is good for the host. **(2)** At the ESS, the retroids have some protection against the accumulation of defective retroids, such as those which produce only templates for example. (It is guaranteed that $W_B < W_A$ in condition (5) and W_B is even smaller than it would be if the retroids were not already engaging in mutual parasitism.) This outcome is good for the retroids.

This latter result is consistent with two strange observations concerning copia to which Kaplan *et al.* (1985) drew attention. (1) They are quite ancient, as inferred from their phylogenetic distribution and (2) they exhibit little variation compared to other transposons. Kaplan *et al.* (1985) put forward a different explanation of these observations. They suggested that each retrotransposon may have exclusive access to its own gene products. This is an interesting suggestion which would set retrotransposons apart from other molecular replicators. However, we are somewhat sceptical as we know that retroviruses do not have such exclusive access.

The mutually parasitic ESS may also be part of the explanation of the odd fact that retrotransposon transcripts may be highly abundant and yet transposition rates are very low. For example, copia transcripts can comprise as much as 3% of the polyadenylated RNA in some cell lines and yet the rate of transposition is less than 10^{-3} per element (Rubin, 1983).

In the discussion of the VSV-DI virus system, we also saw a trade-off between producing templates and producing messenger RNAs. But in the VSV/VSV-DI virus coevolution experiment of De Polo *et al.* (1987), VSV did not evolve to the mutually parasitic ESS just described. This raises the question of what factors channel the coevolution of molecular replicators and their parasites to particular outcomes.

CONCLUSION

In this chapter we have discussed the non-neutral evolutionary biology of molecular replicators, but not their neutral evolution (Kimura, 1983) or evolutionary history. We suspect that interesting questions to be addressed in the near future will be based on a simultaneous consideration of all three aspects. We will not speculate on what these questions might be, but will simply repeat the questions that have arisen in the narrower field covered in this chapter

(1) What are the information requirements of replicase (polymerase) accuracy?

(2) Are small, low-accuracy molecular replicators paying a price in terms of a deleterious mutation load, compared to large, high-accuracy replicators?

(3) Is host jumping, as a virus *strategy*, part of the reason for the short infection period that characterizes low-accuracy viruses?

(4) Can coviruses evolve if only one of the incomplete viruses is a defective interfering virus?

(5) Compared to plant viruses, do animal viruses have lower transmission mutliplicities either between-hosts or between-cells or both?

APPENDIX

The error threshold

Suppose we can divide the molecules in a population into two classes, wild-type and mutant, denoted A and B. The numbers of each are N_A and N_B. The molecules in each class have fitnesses W_A and W_B. For the purpose of this model, fitness may be thought of as the number of copies of each kind of molecule made per unit time. (See Biebricher, Eigen & Gardiner (1985) for a discussion of the meaning of fitness in a particular molecular system.) The molecules have the same death rate, d, per unit time; d may be thought of as a degradation rate. As A is the wild-type and B its defective mutant forms, $W_A > W_B$. These fitnesses and the death rates are not constants; they may change as the molecules become more numerous, for example, But we will suppose that the *ratio* of the fitnesses is constant. A has a replication accuracy of Q, so it mutates to B with probability $1-Q$. We ignore the probability of B mutating to A. This probability is small by com-

parison because there are far more ways of getting something wrong than getting it right–see, for example, the discussion of Fisher's microscope adjustment model given by Kimura (1983).

Our model is

$$\frac{dN_A}{dt} = N_A(QW_A - d)$$
$$\frac{dN_B}{dt} = N_B(W_B - d) + N_A(1-Q)\,W_A. \quad \text{(A 1)}$$

From this model, we can readily derive a necessary condition for the wild-type to exist at equilibrium: it is

$$Q > \frac{W_B}{W_A} \quad \text{(A 2)}$$

So W_B/W_A is a threshold for replication accuracy. If the accuracy, Q, is less than this threshold, then the wild-type cannot persist. The threshold inequality (A 2) can be derived much more directly with a bit of thought. Instead of defining fitness as the number of copies made of a replicator, as is usual in population genetics, we can define fitness as the number of *correct* copies made. With this definition, the fitness of A is QW_A and the fitness of B is W_B. When we bear in mind the mutational flux from A into B, it is clear that a necessary condition for A to persist is

$$QW_A > W_B \quad \text{(A 3)}$$

which can be rearranged to give (A 2). This simple approach was taken by Nee (1987) in his analysis of the evolution of coviruses, which is discussed in the Molecular Games section.

Expression (A 3) implies that, if all mutants are of low fitness (W_B small), then the required accuracy of replication, Q, is also small. The reason for this is as follows. If mutants are almost as fit as the wild-type, they hang about in the population and compete for resources, whereas if all the mutants are lethal the wild-type is free of competition. This fact is highly relevant to the evolution of defective interfering viruses, discussed in the main text. However, it would be wrong to conclude from the model that, if all mutants were lethal, an indefinitely high mutation rate would be permissible, since, obviously, each wild-type must produce at least one wild-type offspring before it dies if the wild-type is to persist. The manifestation of this is the implicit assumption used to derive condition (A 2): we assumed that $QW_A \geqslant d$. If this assumption is not satisfied, the population of molecules becomes extinct.

Model (A 1) is entirely deterministic. If the equilibrium number of wild-type is small, then the wild type may be lost by chance. This is the stochastic process of Muller's Ratchet (Maynard Smith, 1983). The interaction between Muller's Ratchet and Eigen's error threshold has been analysed by Nowak & Schuster (1989).

REFERENCES

BERNS, K. I., MUZYCZKA, N. & HAUSWIRTH, W. W. (1985). Replication of parvoviruses. In *Virology* (ed. Fields, B. N.), pp. 415–32: New York: Raven Press.

BIEBRICHER, C. K., EIGEN, M. & GARDINER, W. C. (1985). Kinetics of RNA replication: competition and selection among self-replicating RNA species. *Biochemistry* **24**, 6550–60.

BOEKE, J. D. (1988). Retrotransposons. In *RNA Genetics*, vol. 2 (ed. Domingo, E., Holland, J. J. & Ahlquist, P.), pp. 59–103. Boca Raton, FL: CRC Press.

BOSCH, V., KUHN, C. & SCHALLER, H. (1988). Hepatitis B virus replication. In *RNA Genetics*, vol. 2 (ed. Domingo, E., Holland, J. J. & Ahlquist, P.), pp. 43–58. Boca Raton, FL: CRC Press.

BROWN, W. M. (1983). Evolution of animal mitochondrial DNA. In *Evolution of Genes and Proteins* (ed. Nei, M. & Koehn, R. K.), pp. 62–88. Sunderland, MA: Sinauer Associates Inc.

BRUENING, G. (1977). Plant covirus systems: two component systems. In *Comprehensive Virology*, vol. 11 (ed. Fraenkel-Conrat, H. & Wagner, R. R.), pp. 55–142. New York: Plenum Press.

BRYAN, G. J., JACOBSON, J. W. & HARTL, D. L. (1987). Heritable somatic excision of a *Drosophila* transposon. *Science* **235**, 1636–8.

CHARLESWORTH, B. (1987). The population biology of transposable elements. *Trends in Ecology and Evolution* **2**, 21–3.

CHARLESWORTH, B. (1988). The maintenance of transposable elements in natural populations. In *Plant Transposable Elements* (ed. Nelson, O. J.), pp. 189–212. New York: Plenum Press.

CHARLESWORTH, B. & LANGLEY, C. H. (1986). The evolution of self-regulated transposition of transposable elements. *Genetics* **112**, 359–83.

COFFIN, J. (1984). Structure of the retroviral genome. In *RNA Tumor Viruses*, 2nd edn, vol. 1 (ed. Weiss, R., Teich, N., Varmus, H. & Coffin, J.), pp. 261–368. New York: Cold Spring Harbor Laboratory.

COFFIN, J. (1988). Replication of retrovirus genomes. In *RNA Genetics*, vol. 2 (ed. Domingo, E., Holland, J. J. & Ahlquist, P.), pp. 3–22. Boca Raton, FL: CRC Press.

CONDIT, R., STEWART, F. M. & LEVIN, B. R. (1988). The population biology of bacterial transposons: a priori conditions for maintenance as parasitic DNA. *American Naturalist* **132**, 129–47.

DEPOLO, N. J., GIACHETTI, C. & HOLLAND, J. J. (1987). Continuing coevolution of virus and defective interfering particles and of viral genome sequences during undiluted passages: virus mutants exhibiting nearly complete resistance to formerly dominant defective interfering particles. *Journal of Virology* **61**, 454–64.

DOMINGO, E. & HOLLAND, J. J. (1988). High error rates, population equilibrium and evolution of RNA replication systems. In *RNA Genetics*, vol. 3 (ed. Domingo, E., Holland, J. J. & Ahlquist, P.), pp. 3–36. Boca Raton, FL: CRC Press

DOOLITTLE, R. F., FENG, D.-F., JOHNSON, M. S. & MCCLURE, M. A. (1989). Origins and evolutionary relationships of retroviruses. *Quarterly Review of Biology* **64**, 1–30.

EIGEN, M. & BIEBRICHER, C. K. (1988). Sequence space and quasispecies distribution. In *RNA Genetics*, vol. 3 (ed. Domingo, E., Holland, J. J. & Ahlquist, P.), pp. 211–45. Boca Raton, FL: CRC Press.

EIGEN, M. & SCHUSTER, M. E. P. (1979). *The Hypercycle*. Berlin: Springer-Verlag.

EMERSON, S. U. (1985). Rhabdoviruses. In *Virology* (ed. Fields, B. N.), pp. 1119–32. New York: Raven Press.

FIELDS, B. N. (1985). *Virology*. New York: Raven Press.

FRANKI, R. I. B. (1985). Plant virus satellites. *Annual Review of Microbiology* **39**, 151–74

FUETTER, J. & HOHN, P. (1987). Involvement of nucleocapsids in reverse transcription: a general phenomenon? *Trends in Biochemical Sciences* **12**, 92–5.

GANEM, D. & VARMUS, H. E. (1987). The molecular biology of the hepatitis B viruses. *Annual Review of Biochemistry* **56**, 651–94.

GOMBOLD, J. L. & RAMIG, R. F. (1986). Analysis of reassortment of genome segments in mice mixedly infected with rotaviruses SA11 and RRV. *Journal of Virology* **57**, 110–16.

HICKEY, D. A. & ROSE, M. R. (1988). The role of gene transfer in the evolution of eukaryotic sex. In *The Evolution of Sex* (ed. Michod, R. E. & Levin, B. R.), pp. 161–75. Sunderland, MA: Sinauer Associates Inc.

HILLMAN, B., CARRINGTON, J. C. & MORRIS, T. J. (1987). A defective interfering RNA that contains a mosaic of a plant virus genome. *Cell* **51**, 427–33.

HOLLAND, J. J. (1985). Generation and replication of defective genomes. In *Virology* (ed. Fields, B. N.), pp. 77–100. New York: Raven Press.

HOLLAND, J. J., SPINDLER, K., HORODYSKI, F., GRABAU, E., NICHOL, S. & VANDEPOL, S. (1982). Rapid evolution of RNA genomes. *Science* **215**, 1577–85.

HUANG, A. S. (1988). Modulation of viral disease processes by defective interfering particles. In *RNA Genetics*, vol. 3 (ed. Domingo, E., Holland, J. J. & Ahlquist, P.), pp. 195–210. Boca Raton FL: CRC Press.

KAPLAN, N., DARDEN, T. & LANGLEY, C. H. (1985). Evolution and extinction of transposable elements in mendelian populations. *Genetics* **109**, 459–80.

KATZ, R. A., KOTLER, M. & SKALKA, A. M. (1988). Cis-acting intron mutations that affect the efficiency of avian retroviral RNA splicing: implication for mechanisms of control. *Journal of Virology* **62**, 2686–95.

KIMURA, M. (1983). *The Neutral Theory of Molecular Evolution*. Oxford: Oxford University Press.

KING, A. M. Q. (1988). Genetic recombination in positive strand RNA viruses In *RNA Genetics*, vol. 2 (ed. Domingo, E. Holland, J. J. & Ahlquist, P.), pp. 149–66. Boca Raton, FL: CRC Press.

KONDRASHOV, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature, London* **336**, 435–40.

KORNBERG, A. (1980). *DNA Replication*. San Francisco: W. H. Freeman & Co.

LAI, M. M. C. (1988). Replication of coronavirus RNA. In *RNA Genetics*, vol. 1 (ed. Domingo, E., Holland, J. J. & Ahlquist, P.), pp. 115–136. Boca Raton, FL: CRC Press.

LEE, T. & NATHANS, D. (1979). Evolutionary variants of SV40: replication and encapsidation of variant DNA. *Virology* **92**, 291–8.

LEVIN, J. G. & ROSENAK, M. J. (1976). Synthesis of murine leukemia virus proteins associated with virions assembled in actinomycin D-treated cells: evidence for persistence of viral messenger RNA. *Proceedings of the National Academy of Sciences, USA* **73**, 1154–8.

LEWIN, B. (1977). *Gene Expression*, vol. 3. New York: John Wiley & Sons.

LINIAL, M. & BLAIR, D. (1984). Genetics of retroviruses. In *RNA Tumor Viruses*, 2nd edn, vol. 1 (ed. Weiss, R., Teich, N., Varmus, H. & Coffin, J.), pp. 649–784. New York: Cold Spring Harbor Laboratory.

LOWY, D. R. (1985). Transformation and oncogenesis: retroviruses. In *Virology* (ed. Fields, B. N.), pp. 235–64. New York: Raven Press.

LURIA, S. E., DARNELL, J. E., BALTIMORE, D. & CAMPBELL, A. (1978). *General Virology*. New York: John Wiley & Sons.

MATTHEWS, R. E. F. (1979). Classification and nomenclature of viruses. *Intervirology* **12**, 150–287.

MAYNARD SMITH, J. (1983). Models of evolution. *Proceedings of the Royal Society of London* B**219**, 315–25.

MAYNARD SMITH, J. (1989). *Evolutionary Genetics*. Oxford: Oxford University Press.

MODRICH, P. (1987). DNA mismatch correction. *Annual Review of Biochemistry* **56**, 435–66.

NEE, S. (1987). The evolution of multicompartmental genomes in viruses. *Journal of Molecular Evolution* **25**, 277–81.

NEE, S. (1989*a*). On the evolution of sex in RNA viruses. *Journal of Theoretical Biology* **138**, 407–12.

NEE, S. (1989*b*). Antagonistic coevolution and the evolution of genotypic randomisation. *Journal of Theoretical Biology* **140**, 499–518.

NOWAK, M. & SCHUSTER, P. (1989). Error thresholds of replication in finite populations. Mutation frequencies and the onset of Muller's Ratchet. *Journal of Theoretical Biology* **137**, 375–95.

O'NEILL, F. J., MARYON, E. B. & CARROLL, D. (1982). Isolation and characterization of defective simian virus 40 genomes which complement for infectivity. *Journal of Virology* **43**, 18–25.

OVERBAUGH, J., DONAHUE, P. R., QUACKENBUSH, S. L., HOOVER, E. A. & MULLINS, J. I. (1988). Molecular cloning of a feline leukemia virus that induces fatal immunodeficiency disease in cats. *Science* **239**, 906–10.

PANGANIBAN, A. T. & FIORE, D. (1988). Ordered interstrand and intrastrand DNA transfer during reverse transcription. *Science* **241**, 1064–9.

PRESSING, J. & REANNEY, D. C. (1984). Divided genomes and intrinsic noise. *Journal of Molecular Evolution* **20**, 135–46.

RE, G. G. & KINGSBURY, D. W. (1988). Paradoxical effects of Sendai virus DI RNA size on survival: inefficient envelopment of small nucleocapsids. *Virology* **165**, 331–7.

REANNEY, D. C. (1982). The evolution of RNA viruses. *Annual Review of Microbiology* **36**, 47–73.

RIESNER, D. & GROSS, H. J. (1985). Viroids. *Annual Review of Biochemistry* **54**, 531–64.

ROIZMANN, B. (1985). Multiplication of viruses: an overview. In *Virology*, (ed. Fields, B. N.), pp. 69–76. New York: Raven Press.

RUBIN, G. M. (1983). Dispersed repetitive DNAs in

Drosophila. In *Mobile Genetic Elements* (ed. Shapiro, J. A.) pp. 329–62. New York: Academic Press.

SCHLESINGER, S. (1988). The generation and amplification of defective interfering RNAs. In *RNA Genetics*, vol. 2 (ed. Domingo, E., Holland, J. J. & Ahlquist, P.), pp. 167–86. Boca Raton, FL: CRC Press.

SHAH, K. V. (1985). Papovaviruses. In *Virology* (ed. Fields, B. N.), pp. 371–93. New York: Raven Press.

SHAPIRO, J. A. (1983). *Mobile Genetic Elements.* New York: Academic Press.

SIMMONS, M. J., RAYMOND, J. D., BOEDIGHEIMER, M. J. & ZUNT, J. R. (1987). The influence of non-autonomous P elements on hybrid dysgenesis in *Drosophila melanogaster*. *Genetics* **117**, 671–85.

SPIEGELMAN, S., MILLS, D. R. & KRAMER, F. R. (1975). The extracellular evolution of structure in replicating RNA molecules. In *Stability and Origin of Biological Information* (ed. Miller, I. R.), New York: Halsted Press.

SPRADLING, A. C. & RUBIN, G. M. (1981). *Drosophila* genome organisation: conserved and dynamic aspects. *Annual Review of Genetics* **15**, 219–64.

STOTT, J. L., OBERST, R. D., CHANNELL, M. B. & OSBURN, B. I. (1987). Genome segment reassortment between two serotypes of bluetongue virus in a natural host. *Journal of Virology* **61**, 2670–80.

SZATHMARY, E. (1989). The integration of the earliest genetic information. *Trends in Ecology and Evolution* **4**, 200–4.

VARMUS, H. & SWANSTROM, R. (1984). Replication of retroviruses. In *RNA Tumor Viruses*, 2nd edn, vol. 1 (ed. Weiss, R., Teich, N., Varmus, H. & Coffin, J.), pp. 369–512. New York: Cold Spring Harbor Laboratory.

WEISSBACH, A. (1977). Eukaryotic DNA polymerases. *Annual Review of Biochemistry* **46**, 25–47.

YOUNGER, J. S., JONES, E. V., KELLY, M. & FRIELLE, D. W. (1981). Generation and amplification of temperature-sensitive mutants during serial undiluted passages of vesicular stomatitis virus. *Virology* **108**, 87–97.

ZIMMERN, D. (1988). Evolution of RNA viruses. In *RNA Genetics*, vol. 2 (ed. Domingo, E., Holland, J. J. & Ahlquist, P.), pp. 211–42. Boca Raton, FL: CRC Press.