Viral quasi-species and recombination

MAARTEN C. BOERLIJST, SEBASTIAN BONHOEFFER AND MARTIN A. NOWAK

Department of Zoology, University of Oxford, Oxford OX1 3PS, U.K.

SUMMARY

Virus populations are complex ensembles of distinct but related genomes (so called quasi-species). Mathematical descriptions of viral quasi-species focus on point mutations as the principal source of variation. However, retroviruses (and many other viruses) are able to recombine their genomes. We study a mathematical model of viral quasi-species dynamics which incorporates both point mutation and recombination. We show that for low mutation rates recombination can reduce the diversity of the quasi-species and enhance overall fitness. For high mutation rates, however, recombination can push the quasi-species over the error threshold, and thereby cause a loss of all genetic information. Finally, recombination introduces bistability to the quasi-species; if the frequency of an advantageous mutant is below a certain threshold, it will not be selected.

1. INTRODUCTION

Quasi-species theory was derived by Eigen, Schuster and coworkers (Eigen 1971; Eigen & Schuster 1979; Eigen *et al.* 1989) to describe the dynamics of replicating nucleic acid molecules under the influence of mutation and selection. The theory was originally developed in the context of pre-biotic evolution of RNA molecules, but in a wider sense it describes any population of reproducing organisms. One of the crucial results of the theory is that selection does not normally lead to a homogeneous population of the fittest type, but instead to an ensemble of genetically distinct variants. This ensemble is called quasi-species.

Virus populations are quasi-species (Domingo *et al.* 1988; Eigen & Biebricher 1988; Holland *et al.* 1992; Wain-Hobson 1992; Duarte *et al.* 1994); they reproduce rapidly and inaccurately, and therefore they usually consist of ensembles of many different mutants. For example, patients infected with the human immunodeficiency virus (HIV) harbour a genetically diverse virus population (Hahn *et al.* 1986; Saag *et al.* 1988; Holmes *et al.* 1992; Wain-Hobson 1993; Cheynier *et al.* 1994). The five hypervariable regions of the HIV envelope protein gp120 have as much as 10–20% intra patient amino acid sequence variation (Holmes *et al.* 1992; Pedroza-Martins *et al.* 1992; Myers *et al.* 1993; Bonhoeffer *et al.* 1995). HIV variation can lead to escape from drug treatment (McLean & Nowak 1992; Larder *et al.* 1993; Frost & McLean 1994; Schuurman *et al.* 1995; Wei *et al.* 1995) and immune responses (Phillips *et al.* 1991) and has been proposed as a main factor of disease progression (Nowak *et al.* 1991; Nowak & McMichael 1995; Nowak *et al.* 1995). Retroviruses – such as HIV – are particularly error prone, because their replication enzyme, the reverse transcriptase, lacks proof reading. In addition retroviruses can recombine their genetic material. Using avian spleen necrosis virus, Temin and coworkers (Hu & Temin 1990; Temin 1991; Temin 1993; Jones *et al.* 1994) have estimated the base substitution rate to be 10^{-5} and the homologous recombination rate to be around 2×10^{-4} (per base pair per replication cycle).

While there is a large literature on models of recombination in diploid and sexually reproducing organisms (Feldman *et al.* 1980; Zhivotovsky *et al.* 1994; Barton 1995), only a few models deal with recombination in an asexual quasi-species (Wiehe 1996). To our knowledge no mathematical model has been proposed that specifically deals with retro-virus recombination. Retroviruses carry two copies of their RNA genome in every virus particle. The replication enzyme, the reverse transcriptase, binds to both RNA molecules simultaneously and produces one DNA molecule. During this act the reverse transcriptase can jump multiple times from one RNA strand to the other, thereby producing a recombinant virus (Hu & Temin 1990). Of course, new recombinant material can only be produced if the two parental strands in the virus particle differ. Such heterozygous virus particles are produced by cells that are simultaneously infected by two different virus strains. Hence, contrary to classical models of recombination, virus recombination depends on superinfection and thus it is density dependent.

In this paper we explore recombination in viral quasi-species. In particular, we study the effects of recombination on the distribution of mutants and the magnitude of the so-called error threshold.

2. VIRAL QUASI-SPECIES: THE ERROR THRESHOLD

Let us first consider viral quasi-species dynamics without recombination. Distinct viral strains are

Proc. R. Soc. Lond. B (1996) **263**, 1577–1584
Printed in Great Britain

represented by bitstrings of length L. The following model describes the change in uninfected cells x, infected cells y_i, and free virus v_i:

$$\frac{dx}{dt} = \lambda - \delta x - x\sum_i \beta_i v_i, \tag{1}$$

$$\frac{dy_i}{dt} = x\sum_j Q_{ij}\beta_j v_j - a_i y_i, \tag{2}$$

$$\frac{dv_i}{dt} = k_i y_i - u_i v_i. \tag{3}$$

In this model λ is the influx rate of uninfected cells; δ, a_i and u_i are the death rates of, respectively, uninfected cells, infected cells, and free virus; β_i is the infection rate; k_i the production rate of new free virus; and Q_{ij} is the probability of strain j mutating to strain i. The mutation matrix is given by:

$$Q_{ij} = p^{H_{ij}}(1-p)^{(L-H_{ij})}. \tag{4}$$

Here p is the mutation rate per bit, and H_{ij} is the Hamming distance between strains i and j, that is the number of bits in which the two strains differ. Error free replication is given by $Q_{ii} = (1-p)^L$.

Without mutation ($p = 0$), the basic reproductive ratio (Diekmann *et al.* 1990; Anderson & May 1991) of viral strain i, is given by $R_i = \lambda\beta_i k_i/(\delta a_i u_i)$. Strains are only viable if $R_i > 1$, and the strain with the largest R_i will outcompete all other strains.

With mutation ($p > 0$) there is a critical error rate, p_c, beyond which the strain with the highest R_i fails to be selected. Let us consider a single peak fitness landscape, where strain F has the highest basic reproductive ratio, R_F, and all other strains have the same, but lower basic reproductive ratio R. Neglecting back mutation the critical error rate is given by $p_c = (R/R_F)^{1/L}$. If $p < p_c$ the quasi-species will be centred around the fittest strain F, which will be most abundant. If $p > p_c$, the fittest strain F will not be selected and each virus strain will have essentially the same relative abundance. This phenomenon is known as the error threshold (Eigen & Schuster 1979; Swetina & Schuster 1982; Eigen *et al.* 1989; Bonhoeffer & Stadler 1993).

In the above virus model there are in fact two different types of error thresholds. If $R < 1$ the virus population will become extinct for $p > p_c$. If $R > 1$ the virus population will survive for $p > p_c$, but the fittest strain will (effectively) disappear. We propose to call this a 'hard' and 'soft' error threshold, respectively. If the mutation rate of the virus population is above the 'soft' error threshold, the virus will survive but the fittest variant will not be selected. The virus population will have a lower average fitness and, therefore, a lower virus load (a lower abundance of virus in the infected individual). As virus load is a main component of disease in many virus infections (Shanley *et al.* 1993; Bangham 1993; Ho 1996; Mellors *et al.* 1996), the mutation rate has a direct influence on viral disease. If the mutation rate is above the 'hard' error threshold, the virus will become extinct. Therefore, increasing the mutation rate of a particular virus with an appropriate drug can in principle reduce virus load and even clear the infection (Eigen 1993).

3. A BASIC PRINCIPLE OF RECOMBINATION

Before we add recombination to the model, we discuss a relation which holds for any type of recombination. Consider two sequences i and j with a genetic distance d_{ij} (for a bitstring model d_{ij} is the Hamming distance). Assume that these sequences recombine to produce an offspring k. If recombination is the only source of variation, we have

$$d_{ik} + d_{jk} = d_{ij}. \tag{5}$$

The genetic difference between the parents equals the sum of the genetic difference between offspring and each of the parents. This relation is important for our understanding of recombination. It shows that in sequence space recombination is always inwards pointing. In the next section this relation will enable us to specify the recombination matrix.

4. BITSTRING RECOMBINATION MODEL

We now adapt equations (1)–(3) to include recombination. We add double infected cells Y_{ij}, which are infected with strain i and super-infected with strain j. v_{ij} is the free virus produced by these super-infected cells, of which 25% will be homozygous type i, 25% will be homozygous type j, and 50% will be heterozygous (assuming Mendelian segregation). The new set of equations becomes:

$$\frac{dx}{dt} = \lambda - \delta x - xV, \tag{6}$$

$$\frac{dy_i}{dt} = xV_i - a_i y_i - sy_i V, \tag{7}$$

$$\frac{dy_{ij}}{dt} = sy_i V_j - a_{ij} y_{ij}, \tag{8}$$

$$\frac{dv_i}{dt} = k_i y_i - u_i v_i, \tag{9}$$

$$\frac{dv_{ij}}{dt} = k_{ij} y_{ij} - u_{ij} v_{ij}. \tag{10}$$

Here s is the rate of super-infection, $V = \sum_i \beta_i v_i + \sum_{ij} \beta_{ij} v_{ij}$ is the sum of all infectious virus, and $V_i = \sum_j Q_{ij}\beta_j v_j + \sum_j Q_{ij} \sum_{kl} M_{jkl}\beta_{kl} v_{kl}$ is the sum of infectious virus of type i, after mutation and recombination, with M_{jkl} being the probability of strain k and l recombining to strain j. All other variables and parameters are as described in equations (1)–(3).

In the case of 'uniform crossover' (that is the recombinant strain has random bits from either parent; Syswerda 1989) the non-zero entries in the recombination matrix M are simply given by:

$$M_{jkl} = 1, \quad \text{if} \quad j = k = l, \tag{11}$$

$$M_{jkl} = \tfrac{1}{2}(1-r) + r(\tfrac{1}{2})^{H_{kl}}, \quad \text{if} \quad j = k \quad \text{or} \quad j = l, \tag{12}$$

$$M_{jkl} = r(\tfrac{1}{2})^{H_{kl}}, \quad \text{if} \quad H_{jk} + H_{jl} = H_{kl}, \tag{13}$$

in which r is the frequency of recombination. This

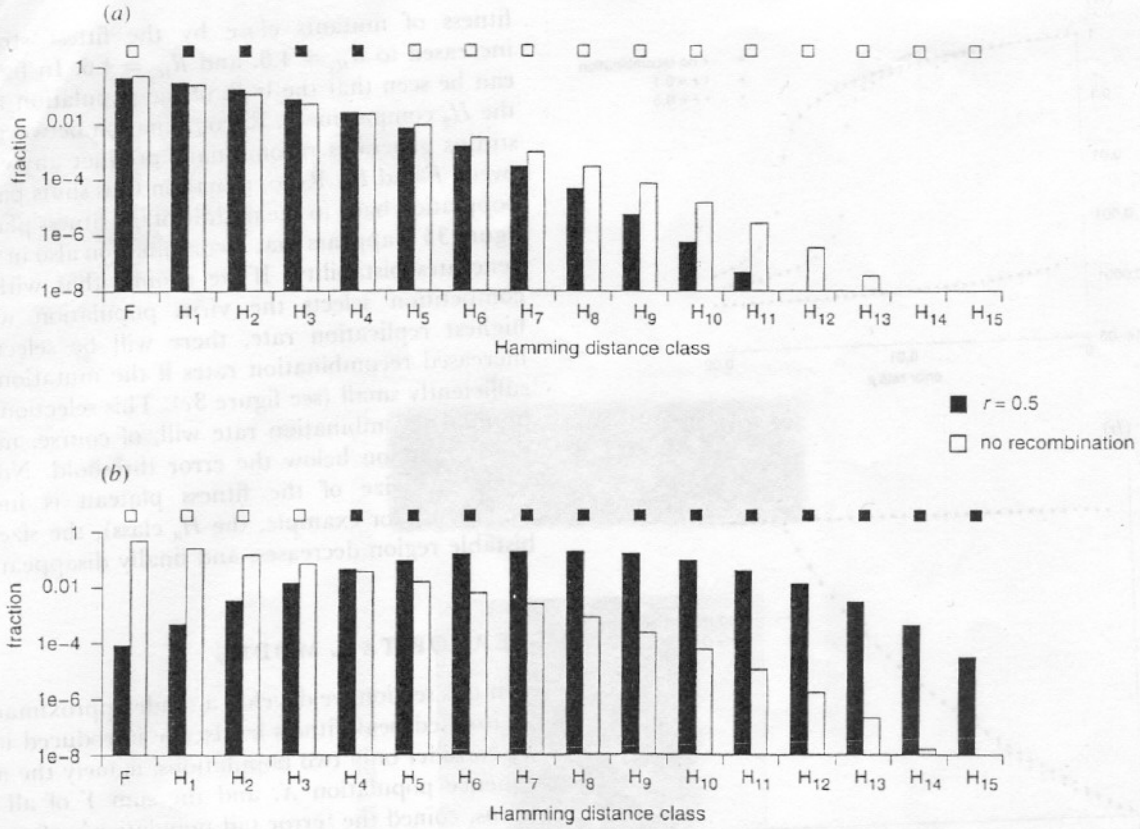

Figure 1. The effect of recombination on quasi-species distribution in an isolated peak fitness landscape. (*a*) Below the error threshold the recombinant population is more compact, because of recombination events involving strain F. But for a slightly increased error rate, in (*b*), it shows that recombination can push the quasi-species over the error threshold. The bulk of the recombinant population is in the H_7 and H_8 compartments, because these contain the most strains. H_i denotes the sum of all mutants with Hamming distance i to F; the small boxes indicate which fraction is larger.

matrix ensures that recombination only acts to reshuffle bits: the total number of zero and one bits in the population is unaffected by recombination. If we use the simplifying assumption that all $\beta_i k_i/u_i$ and $\beta_{ij} k_{ij}/u_{ij}$ are identical, then the steady state structure of equations (6)–(10) is fully determined by the basic reproductive ratios $R_i = \lambda\beta_i k_i/(a_i \delta u_i)$ and $R_{ij} = \lambda\beta_{ij} k_{ij}/(a_{ij}\delta u_{ij})$, the recombination rate r, and the scaled super-infection rate $\rho = s\delta^2 u/(\lambda\beta k)$.

In the next paragraphs we study the steady state structure of equations (6)–(10) for different R_i distributions. We use a default value of $\rho = 0.1$, which means that around 50% of all infected cells are super-infected, and we assume intermediate fitness for super-infected cells, i.e. $R_{ij} = (R_i + R_j)/2$. Bitstrings have length 15. The recombination rate has a maximum at $r = 0.5$, because super-infected cells only produce 50% heterozygote virions, and only heterozygote virions can (effectively) recombine.

(*a*) *Isolated peak landscape*

First we consider the case where only one strain F has an increased R_i value, a so-called 'isolated peak' fitness landscape (Swetina & Schuster 1982). We set $R_F = 5$ and all other $R_i = 3.5$. Figure 1*a* shows the steady state mutant distribution for an error rate of $p = 0.07$, with or without recombination. It turns out that the recombinant population is in some sense more compact: there are less rare mutants, but there is also less of strain F. This effect of recombination can be understood as follows. Most of the population is of strain F. If strain F recombines with e.g. a strain in Hamming distance class 8, then the recombinant product lies anywhere between F and H_8. However, in figure 1*b*, for a slightly increased error rate of $p = 0.08$, recombination drives the population beyond the error threshold, resulting in an almost uniform distribution of mutants. Where recombination acts as a converging operation when F is involved, it acts as a diverging operation in other cases. If for instance two mutants in H_1 recombine, the product lies anywhere between F and H_8. Recombination introduces an instability to the quasi-species composition: it shifts the error threshold, but at low mutation rates it can make the quasi-species more compact.

In figure 2*a* we show the steady rate F population as a function of the error rate. For increasing recombination (i.e. increasing r) the error threshold first sharpens and later (around $r \gtrsim 0.2$) becomes catastrophic, generating a bistable situation where the F dominated population coexist with a mutant dominated population. The recombinant population in figure 1*a* is within this bistable region; there is a second steady state which closely resembles the mutant distribution of the recombinant population in figure

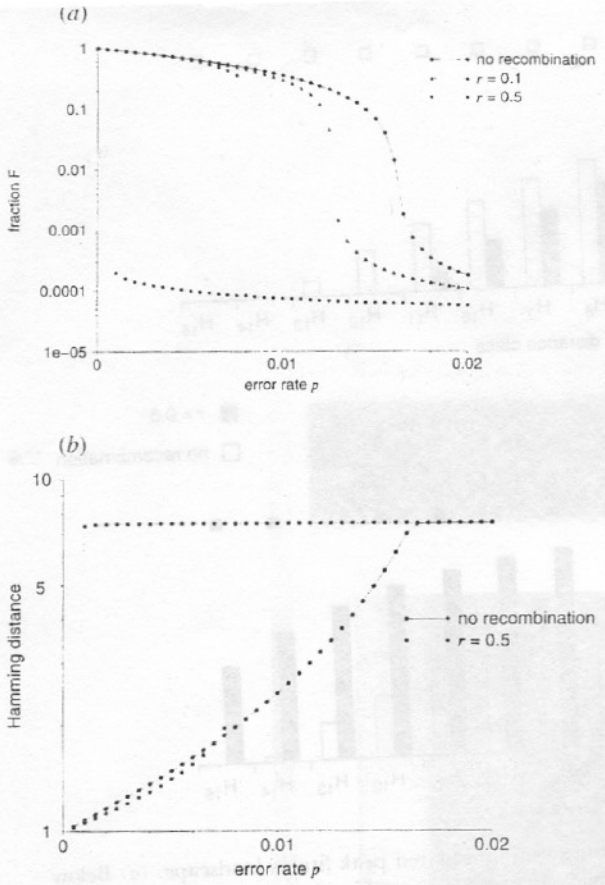

Figure 2. (*a*) Fraction of the fittest strain F as a function of the error rate. For increasing recombination rate the error threshold first becomes steeper, and later on there is a bistable region where the population can maintain itself on the fitness peak, but fails to adapt to it when strain F is not abundant. (*b*) Average Hamming distance of mutants as a function of the error rate. For small mutation rates recombination compacts the mutant population. However, recombination shifts the error threshold towards a lower mutation rate (between $0.0075 \leqslant p \leqslant 0.008$). Beyond the error threshold the average Hamming distance approaches 7.5, which is the average Hamming distance of random strains of length 15.

1*b*. An intuitive explanation for this bistability is that at high frequency of the F strain recombination acts to compact the quasi-species, whereas at low frequency of F recombination acts to randomize the quasi-species. In §5 we show that generating bistability is a fundamental property of recombination. In figure 2*b* we plot the average Hamming distance (to the F strain) of mutants as a function of the error rate. Recombination decreases diversity for small mutation rates, but it shifts the error threshold and creates bistability.

(*b*) *Plateau landscape*

In the isolated peak landscape, recombination is always disadvantageous for the virus, because it decreases the abundance of F and it shifts the error threshold towards lower mutation rates. Recombination, however, can be advantageous in the case of correlated fitness landscapes (see, for example, Kondrashov 1988). Consider a situation, where the fitness of mutants close by the fittest strain F is increased to $R_{H_1} = 4.8$, and $R_{H_2} = 4.6$. In figure 3*a* it can be seen that the bulk of the population now is in the H_2 compartment. Recombination between two H_2 strains generates recombinant product anywhere between F and H_4. Recombination thus shifts part of the population back to the middle of the fitness plateau. In figure 3*b* it appears that recombination also in this case generates bistability. If we assume that within host competition selects the virus population with the highest replication rate, there will be selection for increased recombination rates if the mutation rate is sufficiently small (see figure 3*c*). This selection for the optimal recombination rate will, of course, maintain the population below the error threshold. Note that when the size of the fitness plateau is increased (including, for example, the H_3 class), the size of the bistable region decreases, and finally disappears.

5. ERROR TAIL MODEL

In this section we develop a crude approximation of the isolated peak fitness landscape introduced in §4*a*. We consider only two populations, namely the master sequence population X, and the sum Y of all other strains, coined the 'error tail population' (after Eigen & Schuster 1979). X and Y are given in frequencies, so $Y = 1 - X$. Both the master sequence and the error tail can replicate, but the error tail replicates at a reduced rate $1 - \sigma$, where σ is the selection coefficient (Crow & Kimura 1970). The master sequence mutates with a rate p into the error tail, 'back mutations' are neglected. In the model without recombination the change in the frequency of X is given by:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = (1-p)\,X - (X + (1-\sigma)\,Y)\,X. \qquad (14)$$

In this model the error threshold is at $p = \sigma$, i.e. for error rates that exceed σ the master sequence will disappear completely. For $p < \sigma$ there is a steady state at $X = 1 - p/\sigma$. The $X' = 0$ isocline as a function of p is shown in figure 4*a*.

To this model we add recombination at rate r. As with mutation, we neglect 'back recombination' of two error tail mutants to the master sequence. Therefore, we only have to consider recombination between a master sequence and a mutant. If the error tail population is small, mutants will in the limit only have one mutation, and recombination is a neutral operation. If on the other hand the error tail population is large, mutants will have many mutations, and the product of recombination between a master sequence and a mutant is (almost) always a mutant. For simplicity we assume a linear relation between these two cases, i.e. $f(X) = 1 - X$ in equation (15). The equation with recombination thus becomes:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = (1-p)\,X - rXYf(X) - (X + (1-\sigma)\,Y)\,X, \qquad (15)$$

figure 4*b–d* show the three distinct cases for the $X' = 0$ isoclines as a function of the error rate p. For $r \leqslant \sigma/2$,

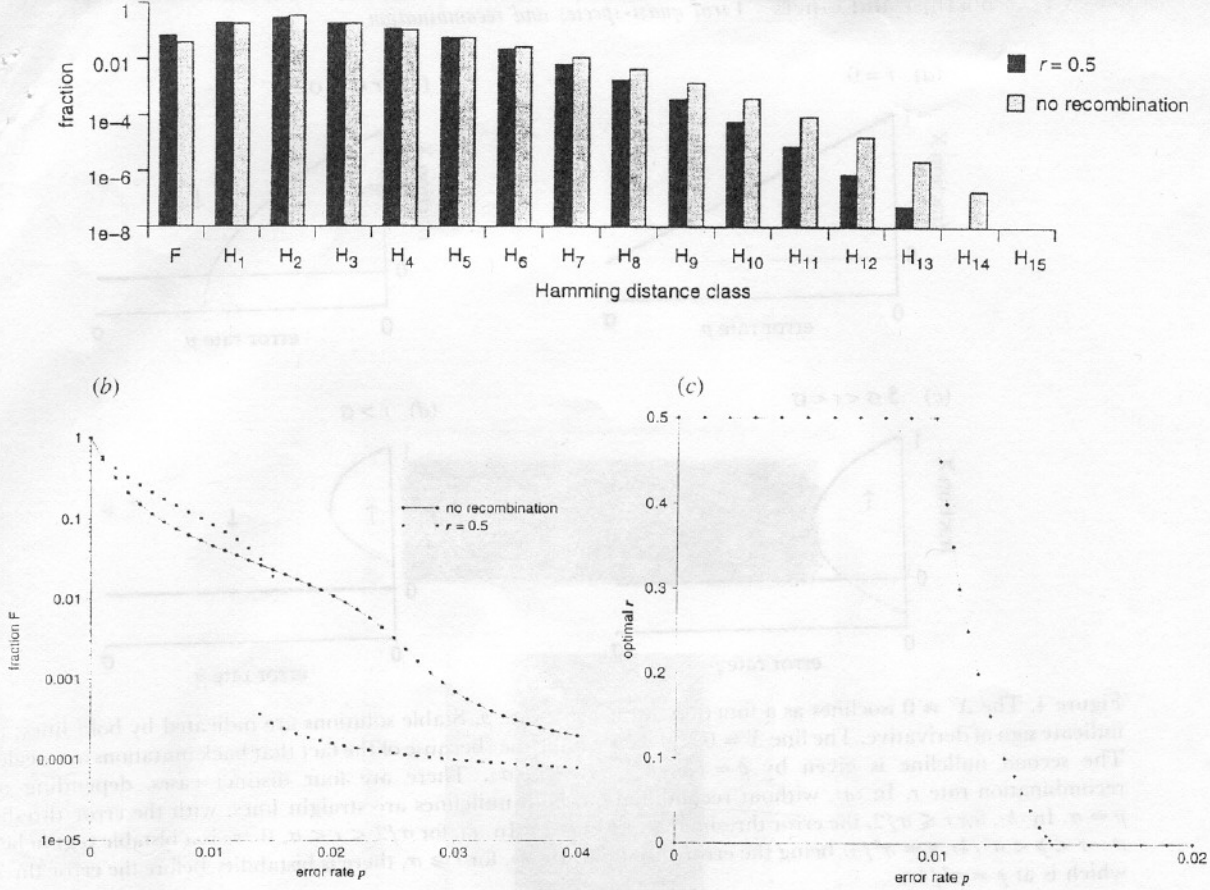

Figure 3. Recombination in a plateau fitness landscape. Now the strains in the H_1 and H_2 compartment also have an increased fitness (see text for parameter values). In (*a*), before the error threshold at $p = 0.011$, it shows that the recombinant population is again more compact, and it has more of its mass in the middle of the fitness plateau. In (*b*) it appears that recombination also in this case induces bistability, with the error threshold at the point where recombination can no longer keep the population in the middle of the fitness plateau. The error threshold without recombination is in this case very smooth, acting around $p = 0.05$. In (*c*) it shows that recombination is advantageous to the virus for small mutation rates (we assume $r \leqslant 0.5$).

in figure 4*b*, there is a 'normal' error threshold. For $r > \sigma/2$, in figures 4*c–d*, there is a 'catastrophic' error threshold. For $\sigma/2 < r < \sigma$, in figure 4*c*, there is bistability for error rates between $\sigma - r \leqslant p \leqslant \sigma^2/4r$, i.e. there is an 'accumulation threshold' below which the master sequence cannot establish itself. If $r \geqslant \sigma$, in figure 4*d*, for error rates smaller than the error threshold $p < \sigma^2/4r$ there is always an accumulation threshold. Thus, bistability arises if the recombination rate is large compared with the selection coefficient, and recombination shifts the error threshold to lower mutation rates.

6. DISCUSSION

In this paper we demonstrated several implications of recombination for viral quasi-species. First, for small mutation rates (i.e. below the error threshold), recombination can focus the quasi-species around a fitness optimum. In this sense, recombination acts as an error repair mechanism (Temin 1991), but it also means that the population is less flexible to environmental change. Recombination introduces hysteresis to the quasi-species: the population does not necessarily adapt to the highest fitness peak, but instead it favours the peak where it currently resides. We have shown that, in the case of an isolated fitness peak or a small fitness plateau, there can even be bistability between the fittest mutant and the error tail. We confirmed this result in the mini model in §5; the bistability depends on the amount of recombination relative to the selection coefficient. Furthermore, recombination shifts the error threshold to lower mutation rates. Near the error threshold, without recombination, the fittest strain only makes up a small percentage of the total population (Eigen *et al.* 1989). Under such conditions recombination acts as a diverging operation, driving the population beyond the error threshold. There can be selection for recombination if fitness is correlated and if the mutation rate is sufficiently small (as shown in §4*b*). Again, this selection for recombination will bring the population closer to the error threshold (but not beyond it).

We have extensively tested the diploid bitstring model for other fitness distributions, such as 'smooth' fitness peaks with additive or multiplicative contri-

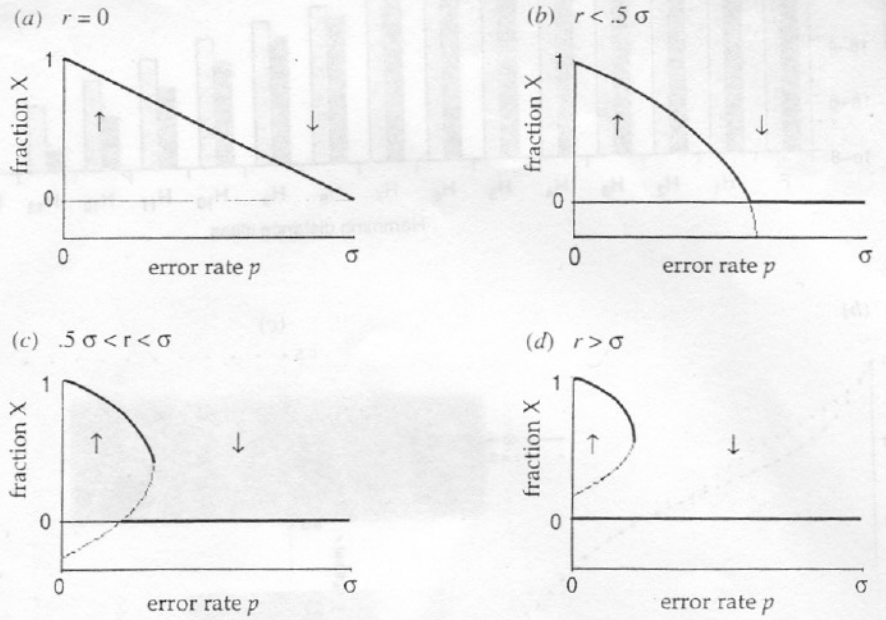

Figure 4. The $X' = 0$ isoclines as a function of the error rate p. Stable solutions are indicated by bold lines; arrows indicate sign of derivative. The line $X = 0$ is always a nullcline (because of the fact that back mutations are neglected). The second nullcline is given by $p = (X-1)(-rX+r-\sigma)$. There are four distinct cases, depending on the recombination rate r. In (*a*), without recombination, both nullclines are straight lines, with the error threshold at $p = \sigma$. In (*b*), for $r \leqslant \sigma/2$, the error threshold is at $p = \sigma - r$. In (*c*), for $\sigma/2 < r < \sigma$, there is a bistable region between $\sigma - r < p < \sigma^2/4r$, $p = \sigma^2/4r$ being the error threshold. In (*d*), for $r \geqslant \sigma$, there is bistability before the error threshold, which is at $p = \sigma^2/4r$.

butions of the distinct bits, multiple peaks and random distributions. We varied the super-infection rate ρ, and the β_i, k_i, u_i, and $\beta_{ij}, k_{ij}, u_{ij}$ parameters. We have, furthermore, looked at alternatives to uniform crossover, such as one-point crossover and multiple crossover with a fixed crossover probability. We also tested a haploid model, where no assumptions about heterozygote fitness have to be made. In all these cases the main conclusion holds: recombination shifts the error threshold towards lower mutation rates and it makes the transition sharper. The effect intensifies as fitness correlations are more localized, and it accumulates in bistability (or multiple stability in the case of multiple fitness peaks). The same type of bistability has been demonstrated in a two locus two allele model of quasi-species dynamics which neglects back mutations (Wiehe 1996). A similar effect has been observed in a diploid sexual model with recombination (Higgs 1994), where the error threshold first sharpens and later becomes discontinuous for an increasing dominance coefficient (i.e. decreasing heterozygote fitness). However, bistability also appears in diploid models without recombination for low heterozygote fitness (Nagylaki 1992; Wiehe *et al.* 1995; Baake & Wiehe 1996). We conclude that recombination and low heterozygote fitness are separate mechanisms which can both generate bistability.

Our results apply to steady state analysis on fixed fitness landscapes. This, of course, is only a first step in understanding the effects of recombination on viral quasi-species. An important difference between sexual recombination and viral recombination is that the latter is density dependent, and thus recombination events become more frequent as the viral population size increases, for example in the later stages of an HIV infection. Furthermore, the fitness landscape might be changing during an infection, for example, caused by varying immune responses (Bonhoeffer *et al.* 1995). Recombination might play a role in preventing extinction of advantageous mutants during selective sweeps (Barton 1995; Wiehe 1996). Recombination can account for large jumps in sequence space if a patient becomes multiply infected with virus of divergent strains (Robertson *et al.* 1995). Recombination can act as an important repair mechanism to overcome 'breaks' in the RNA molecule (Temin 1991; Coffin 1992). Finally, virus populations are finite and initially small. Recombination can prevent extinction of advantageous mutants because of random drift (Nowak & Schuster 1989; Otto & Barton 1996).

We thank D. C. Krakauer and an anonymous reviewer for helpful discussion and comments on the manuscript. M. C. B. is funded by E.C. grant number ERBCHBICT941834. S. B. and M. A. M. are funded by the Wellcome Trust.

REFERENCES

Anderson, R. M. & May, R. M. 1991 *Infectious diseases of humans*. Oxford University Press.

Baake, E. & Wiehe, T. 1996 Bifurcations in haploid and diploid sequence space models. *Math. Biol.* (In the press.)

Bangham, C. R. M. 1993 The influence of immunity on virus evolution. In *The molecular evolution of viruses* (ed. A. Gibbs, C. Calisher & F. Garcia-Arenal). Cambridge University Press.

Barton, N. H. 1995 A general model for the evolution of recombination. *Genet. Res.* **65**, 123–144.

Bonhoeffer, S., Holmes, E. C. & Nowak, M. A. 1995 Causes of HIV diversity. *Nature, Lond.* **376**, 125.

Bonhoeffer, S. & Stadler, P. F. 1993 Error thresholds on complex value landscapes. *J. theor. Biol.* **162**, 359–372.

Cheynier, R., Henrichwark, S., Hadida, F. *et al.* 1994 HIV and T-cell expansion in splenic white pulps is accompanied by infiltration of HIV-specific cytotoxic T lymphocytes. *Cell* **12**, 373–387.

Coffin, J. 1992 Genetic diversity and evolution of retroviruses. *Curr. Top. Microbiol. Immunol.* **176**, 143–164.

Crow, J. F. & Kimura, M. 1970 *An introduction to population genetics theory*. New York: Harper and Row.

Diekmann, O., Heesterbeek, J. A. P. & Metz, J. A. J. 1990 On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382.

Domingo, E., Holland, J. J. & Ahlquist, P. 1988 *RNA Genetics* volume 1–3. Florida: CRC press.

Duarte, E. A., Novella, I. S., Weaver, S. C. *et al.* 1994 RNA virus quasispecies: significance for viral disease and epidemiology. *Infect. Agents Dis.* **3**, 201–214.

Eigen, M. 1971 Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **10**, 465–523.

Eigen, M. 1993 Viral quasispecies. *Sci. Am.* **269**, 32–39.

Eigen, M. & Biebricher, C. 1988 Sequence space and quasispecies distribution. In *RNA genetics* (ed. E. Domingo, J. J. Holland & P. Ahlquist), pp. 211–245. Florida: CRC press.

Eigen, M., McCaskill, J. & Schuster, P. 1989 The molecular quasi-species. *Adv. Chem. Phys.* **75**, 149–263.

Eigen, M. & Schuster, P. 1979 *The hypercycle – a principle of natural selforganization*. Berlin: Springer.

Feldman, M. W., Christiansen, F. B. & Brooks, L. D. 1980 Evolution of recombination in a constant environment. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4838–4841.

Frost, S. D. W. & McLean, A. R. 1994 Quasispecies dynamics and the emergence of drug resistance during zidovudine therapy of hiv infection. *AIDS* **8**, 323–332.

Hahn, B. H., Shaw, G. M., Taylor, M. E. *et al.* 1986 Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. *Science, Wash.* **232**, 1548–1553.

Higgs, P. G. 1994 Error thresholds and stationary mutant distributions in multi-locus diploid genetic models. *Genet. Res., Camb.* **63**, 63–78.

Ho, D. D. 1996 Viral counts in HIV infection. *Science, Wash.* **272**, 1124–1125.

Holland, J., Torre, J. D. L. & Steinhauer, D. 1992 RNA virus populations as quasispecies. *Curr. Top. Microbiol. Immunol.* **176**, 1–20.

Holmes, E., Zhang, L., Simmonds, P., Ludlam, C. & Brown, A. L. 1992 Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. *Proc. natn. Acad. Sci. U.S.A.* **89**, 4835–4839.

Hu, W. S. & Temin, H. W. 1990 Retroviral recombination and reverse transcription. *Science, Wash.* **250**, 1227–1233.

Jones, J. S., Allan, R. W. & Temin, H. W. 1994 One retroviral RNA is sufficient for synthesis of viral DNA. *J. Virol.* **68**, 207–216.

Kondrashov, A. 1988 Deleterious mutations and the evolution of sexual reproduction. *Nature, Lond.* **336**, 435–440.

Larder, B., Kohli, A., Kellam, P., Kemp, S., Kronick, M. & Henfrey, R. 1993 Quantitative detection of HIV-1 drug resistance mutations by automated DNA sequencing. *Nature, Lond.* **365**, 671–673.

McLean, A. & Nowak, M. 1992 Competition between zidovudine-sensitive and zidovudine-resistant strains of HIV. *AIDS* **6**, 71–79.

Mellors, J. W., Rinaldo Jr, C. R., Gupta, P., White, R. M., Todd, J. A. & Kingsley, L. A. 1996 Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science, Wash.* **272**, 1167–1170.

Myers, G. K., Korber, B., Berzofsky, J. A. & Smith, R. F. 1993 *Human retroviruses and AIDS 1993*. New Mexico: Los Alamos National Laboratory.

Nagylaki, T. 1992 *Introduction to theoretical population genetics*. New York: Springer.

Nowak, M. & Schuster, P. 1989 Error thresholds of replication in finite populations, mutation frequencies and the onset of Muller's ratchet. *J. theor. Biol.* **137**, 375–395.

Nowak, M. A., Anderson, R. M., McLean, A. R., Wolfs, T. F., Goudsmit, J. & May, R. M. 1991 Antigenic diversity thresholds and the development of AIDS. *Science, Wash.* **254**, 963–969.

Nowak, M. A., May, R. M., Philips, R. E. *et al.* 1995 Antigenic oscillations and shifting immunodominance in HIV-1 infections. *Nature, Lond.* **375**, 606–611.

Nowak, M. A. & McMichael, A. J. 1995 How HIV defeats the immune system. *Sci. Am.* **273**, 58–65.

Otto, S. P. & Barton, N. H. 1996 The evolution of recombination: removing the limits of natural selection. (Submitted.)

Pedroza-Martins, L., Chenciner, N. & Wain-Hobson, S. 1992 Complex intrapatient sequence variation in the V1 and V2 hypervariable regions of the HIV-1 gp120 envelope sequence. *Virology* **191**, 837–845.

Phillips, R. E., Rowland-Jones, S., Nixon, D. F. *et al.* 1991 Human immunodeficiency virus genetic variation that can escape cytotoxic t cell recognition. *Nature, Lond.* **354**, 453–459.

Robertson, D. L., Sharp, P. M., McCutchan, F. E. & Hahn, B. H. 1995 Recombination in HIV-1. *Nature, Lond.* **374**, 124–126.

Saag, M. S., Hahn, B. H., Gibbons, J., Li, Y., Parks, E. S., Parks, W. P. & Shaw, G. M. 1988 Extensive variation of human immunodeficiency virus type-1 *in vivo*. *Nature, Lond.* **334**, 440–444.

Schuurman, R., Nijhuis, M., van Leeuwen, R. *et al.* 1995 Rapid changes in human immunodeficiency virus type 1 RNA load and appearance of drug-resistant virus populations in persons treated with lamivudine (3TC). *J. Infect. Dis.* **171**, 1411–1419.

Shanley, J. D., Biczak, L. & Forman, S. J. 1993 Acute murine cytomegalovirus infection induces lethal hepatitis. *J. Infect. Dis.* **167**, 264–269.

Swetina, J. & Schuster, P. 1982 Self-Replication with errors, a model for polynucleotide replication. *Biophys. Chem.* **16**, 329–345.

Syswerda, G. 1989 Uniform crossover in genetic algorithms. In *Proceedings of the third international conference on genetic algorithms and their applications* (ed. J. D. Schaffer), pp. 2–9. Morgan Kaufmann.

Temin, H. M. 1991 Sex and recombination in retroviruses. *Trends Genet.* **7**, 71–74.

Temin, H. M. 1993 Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proc. natn. Acad. Sci. U.S.A.* **90**, 6900–6903.

Wain-Hobson, S. 1992 Human immunodeficiency virus type 1 quasispecies *in vivo* and *ex vivo*. *Curr. Top. Microbiol. Immunol.* **176**, 181–193.

Wain-Hobson, S. 1993 The fastest genome evolution ever

scribed: HIV variation *in situ*. *Curr. Opin. Genet. Dev.* **3**, 8–883.

X., Gosh, S. K., Taylor, M. E. *et al.* 1995 Viral namics in human immunodeficiency virus type 1 ection. *Nature, Lond.* **373**, 117–122.

e, T. 1996 Two locus recombination produces opposite ects which depend on the mode of selection. (Subtted.)

e, T., Baake, E. & Schuster, P. 1995 Error propagation in reproduction of diploid organisms. *J. theor. Biol.* **177**, 1–15.

Zhivotovsky, L. A., Feldman, M. W. & Christiansen, F. B. 1994 Evolution of recombination among multiple selected loci: a generalized reduction principle. *Proc. natn. Acad. Sci. U.S.A.* **91**, 1079–1083.

Received 15 July 1996; accepted 14 August 1996