ORIGIN OF SEX REVISITED

MAURO SANTOS[1,2*], ELIAS ZINTZARAS[1,3] and EÖRS SZATHMÁRY[1,4]

[1] *Collegium Budapest, Institute for Advanced Study, Szentháromság u. 2, Budapest, Hungary;*
[2] *Departament de Genètica i de Microbiologia, Grup de Biologia Evolutiva, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain;* [3] *National Agricultural Research Foundation (N.AG.RE.F.), 19 Egialias and Chalepa, Maroussi, Greece;* [4] *Department of Plant Taxonomy and Ecology, Eötvös University, 2 Ludovika tér, Budapest, Hungary*
(* author for correspondence, e-mail: mauro.santos@uab.es)

Abstract. Why did sex ever arise in the first place? Why it does not disappear in view of the greater efficiency of asexuals? These are clearly two different questions, and we suggest here that the solution for the origin of sex does not necessarily come from theoretical considerations based on currently existing genetic systems. Thus, while we agree with a number of authors in that the emergence of sex (understood as the exchange of genetic material between genomes) is deeply rooted in the origin of life and happened during the very early stages in the transition from individual genes ('replicators') to bacteria-like cells ('reproducers'), we challenge the idea that recombinational repair was the major selective force for the emergence of sex. Taking the stochastic corrector model as a starting point, we provide arguments that question the putative costs of redundancy in primitive protocells. In addition, if genes that cause intragenomic conflict (i.e., parasites) are taken into account, it is certainly wrong to suggest that cellular fusion would be beneficial at the population level (although this strong claim needs some qualifications). However, when a continuous input of deleterious mutations that impair the fitness of the protocell as a whole is considered in the model (in the realistic range in which stable mutant distributions of quasi-species within compartments are established), there are circumstances when sex could be beneficial as a side effect of the dynamic equilibrium between cellular fusion-mutation-selection. The scenario we have explored numerically is fully consistent with the idea that the universal ancestor was not a discrete entity but an ensemble of proto-organisms that exchanged much genetic information.

Keywords: competition, cooperation, deleterious mutations, genomic conflict, origin of sex, parasites, selfish replicators, stochastic corrector model

1. Introduction: Did Sex First Evolve to Repair Genetic Damage?

The continued persistence of sex, understood as the exchange of genetic material between genomes (Michod and Levin, 1988; but see below), is puzzling in view of the expected two-fold advantage of asexuals (Williams, 1975; Maynard Smith, 1978; Hurst and Peck, 1996). It has been suggested that sexual reproduction is maintained by selection against recurrent deleterious mutations, although recent studies point to a combination of different mechanisms being responsible for its continued persistence (West *et al.*, 1999; Keightley and Eyre-Walker, 2000). But, why did sex ever arise in the first place? This is a different question from that

of its maintenance. The necessity to repair genetic damage has been claimed to be the immediate factor responsible for the origin of sex (Bernstein *et al.*, 1984, 1985; Michod, 1993). It is true that diploids are more resistant to damaging agents than haploids, and for non-phagotrophs it seems to hold (with exceptions) that haploid organisms may have a growth advantage because of its smaller cell size (see Cavalier-Smith, 1985; Maynard Smith and Szathmáry, 1995). However, a problem with the repair theory is that the origin and maintenance of a haploid/diploid cycle can be solely explained by a faster removal in haploids of recurrent deleterious mutations that are partially expressed in diploids (Kondrashov and Crow, 1991; Perrot *et al.*, 1991). Yet, there is a more fundamental snag with the idea that the origin of sex is based on overcoming genetic damage; namely, to assume as Bernstein *et al.* (1984) did that gene redundancy was already costly in the first protobionts.

Under the standard genes-first view of the origin of life (Gilbert, 1986; Joyce *et al.*, 1987; Gesteland *et al.*, 1999) it is supposed that at some time the naked self-replicating oligonucleotide analogues (RNA-like molecules) managed to clothe themselves in a cell-like structure (protocell) that enclosed either a cyclically coupled system of autocatalytic and cross-catalytic molecular mutualists (i.e., a hypercycle; see Eigen and Schuster, 1979; Eigen *et al.*, 1981), or a non-hypercyclic system of unlinked competing genes replicated by a non-specific replicase (whose dynamics is described by the 'stochastic corrector model'; see Szathmáry and Demeter, 1987; Grey *et al.*, 1995). Those first genes might initially have coded for almost nothing except the ability to make copies of themselves with an extremely high mutation rate (Friedberg *et al.*, 1995; see Johnston *et al.*, 2001), and before the evolution of transcription both genes and enzymes were plus- and minus-strand of RNA-like templates that would have about equal concentrations inside compartments (see below). RNA-based replication and catalysis is hence assumed to have originated before protein synthesis became a major biochemical pathway. Definitely protocells would not have possessed machinery for accurately segregating gene copies, and irregular reduction by random distribution of genes between two daughter cells was the most likely mechanism. Consequently, they would have needed to contain redundant copies of each kind of gene so that the probability of transmission of at least some copies to each daughter protocell after replication and stochastic fission would be large enough for positive population growth (Niesert *et al.*, 1981; Koch, 1984; Reanney, 1987).

Contrary to the claims by Bernstein *et al.* (1984) redundancy is costly (as in organisms today) only when the necessary stoichiometric relationships between gene products is unbalanced, which may be maintained by stabilizing selection. Thus, current evidence suggests that selection has acted largely to silence duplicated genes and reduce the rate of gene duplication because duplicates of single genes are out of balance with their interacting partners (Lynch and Conery, 2000; Otto and Yong, 2002), which agrees with the numerous observations that gene duplications that increase protein dosage can be pathogenic (e.g., Lupski *et al.*,

1996; Mergenthaler *et al.*, 2001; McDermid and Morrow, 2002). Those gene duplications that persist in an evolving lineage might originally have a net advantage due primarily to a protein dosage effect in response to variable environmental conditions (Kondrashov *et al.*, 2002). The situation with single duplicate genes, however, markedly contrasts with the high level of duplicate-gene preservation observed after polyploidy and the potential for increased adaptability in polyploid lineages (Amores *et al.*, 1998; Cronn *et al.*, 1999; Otto and Whitton, 2000).

Before enzymes and genes became replicationally uncoupled both the + and − strands of RNA-like templates must be equipped with a recognition site ('target') for the replicase. Assuming that replication goes in the $5' \rightarrow 3'$ direction as today, the $3'$ and $5'$ ends of the same strand must therefore be complementary. To the extent that the + and the − strands possessed identical targets, there should be a complete symmetry in replication rates and, hence, an approximate stoichiometric proportion in the concentration of both strands (Eigen, 1971). Redundancy could then be preserved at no cost and, in theory the number of copies of each gene (ribozyme) in primitive protocells could have been extremely high. As suggested by Koch (1984), the potential upper limit to gene copy number was set by the risk that Darwinian selection would be stopped because of dilution of favourable mutations in an 'orgy of redundancy'. Under this scenario it is easy to envisage that the putative benefits of periodical fusion between protocells to overcome the problems created by genetic damage were negligible in comparison with the real costs of horizontal gene transfer of selfish mutants (parasites). In their original work Bernstein *et al.* (1984) assumed no genetic variability in already cooperative biochemical symbionts that enhanced the replication rate of their partners (i.e., a hypercycle). Therefore, they neglected the potential danger of parasites spreading in the population as a result of cellular fusion (or migration among lineages), which creates an opportunity for the evolution of genetic conflicts (Werren *et al.*, 1988; Partridge and Hurst, 1998) and could eventually destroy the hypercycle (see Maynard Smith, 1979; Bresch *et al.*, 1980).

Here we critically analyse the effects resulting from periodical fusion of protocells before division by binary fission on average population fitness. We assume that primitive protocells enclosed a non-hypercyclic system of competing genes replicated by a non-specific replicase (i.e., we take as a starting point the stochastic corrector model). Recent simulations suggest that such a population is an efficient information integrator system and can tolerate higher deleterious mutation rates (i.e., reaches a lower equilibrium mutational load) than a population of protocells enclosing hypercycles (Zintzaras *et al.*, 2002). Multiple copies of each gene were obviously needed to ensure a large enough probability of transmission to daughter protocells because gene segregation was not yet accurate (Figure 1). Differential growth of replicators ('gene selection') would lead to deterioration of compartments, but selection on stochastically produced offspring variants ('between-protocell selection') could rescue the population from extinction (favouring cooperative molecules).
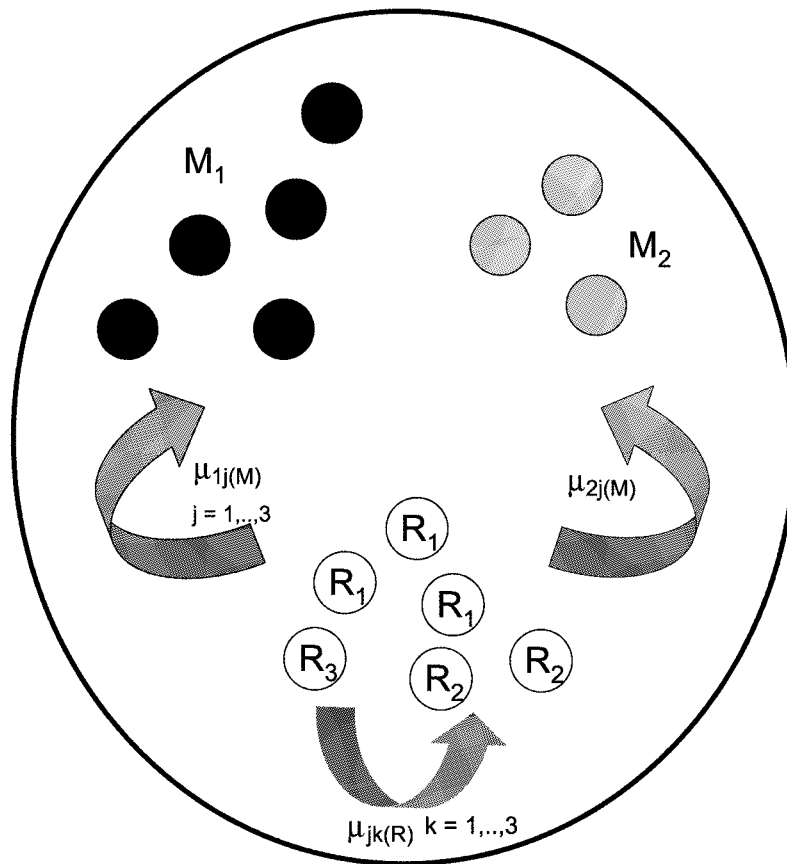
Figure 1. Schematic representation of a protocell with two different templates that stand for metabolic genes (black circles are M_1, grey circles are M_2) essential for growth and survival that are replicated by genes (R_j; open circles) with a replicase function. The templates are free to compete within the compartment and, therefore, may have different replication rates (μ's). As a consequence of cellular fusion (sex), different R_j templates (i.e., derived from varied ancestors) can coexist within a given protocell.

2. The Models

Two kinds of mutations are worth considering: (i) a mutated gene that acquires an increase target affinity toward the replicase and outperforms its partners within the protocell (i.e., a 'selfish' molecule); and (ii) deleterious mutations that impair the fitness of the protocell as a whole.

2.1. NO DELETERIOUS MUTATIONS

We assume here that each protocell has three kinds of 'wild-type' templates: a non-specific replicase (R_j, $j = 1, \ldots, m$; see below) and two target metabolic genes

(M_1, M_2; *sensu* Gánti, 1975, 2003) essential for growth and survivorship (Figure 1). For a particular protocell, its fitness is 1 if at least one copy of each template's class is present; otherwise the fitness is 0 and the cell is set as dead. Thus, the average fitness of the population at any generation is simply the proportion of protocells that remain alive. The Monte Carlo model is shown in Figure 2. A generation starts with a population of K (set to 500) protocells with n templates (genes) initially at equal concentrations (i.e., $R_j = n/3$, $R_k = 0, k \neq j$; $M_1 = M_2 = n/3$ at t_0). A protocell is randomly chosen according to its relative fitness for template replication. A template is chosen randomly and is replicated according to its probability of replication (see below). In the stochastic formulation it is assumed that protocells divide when reaching a critical size, which is defined here as the doubling of the total number of genes. Thus, if the number of genes after template replication is less than $2n$ the protocell is turned back to the population; otherwise the cell divides by randomly assorting the templates to two daughter cells (i.e., we follow the more realistic 'continuous version' of the stochastic corrector model; see Zintzaras *et al.*, 2002). The procedure continues until the population size increases to $2K$, then half of the cells are discarded at random and the start of a new generation is assumed.

The rate of replication (μ) of a given template within protocells depends on two factors: its target affinity toward the replicase and the travelling speed of the replicase along the template (i.e., the replicase activity). Because the replicase and the target are two physically independent molecules, random sampling of both replicase and target template is done without replacement. The differential equations governing the dynamics of templates' growth before cell division are:

$$\frac{d(M_i \mid R_j)}{dt} = \mu_{ij(M)} M_i R_j, \quad i = 1, 2, j = 1, \ldots, m , \tag{1}$$

for the replication of target metabolic templates, and

$$\frac{d(R_j \mid R_k)}{dt} = \mu_{jk(R)} R_j R_k, \quad k = 1, \ldots, m , \tag{2}$$

for the replication of target replicase. The probabilities of replicating a template depend on the previous growth rates and are calculated as follows:

$$P(M_i \mid R_j) = \frac{\dfrac{d(M_i \mid R_j)}{dt}}{\displaystyle\sum_i \sum_j \frac{d(M_i \mid R_j)}{dt} + \sum_j \sum_k \frac{d(R_j \mid R_k)}{dt}} ,$$

$$P(R_j \mid R_k) = \frac{\dfrac{d(R_j \mid R_k)}{dt}}{\displaystyle\sum_i \sum_j \frac{d(M_i \mid R_j)}{dt} + \sum_j \sum_k \frac{d(R_j \mid R_k)}{dt}} . \tag{3}$$
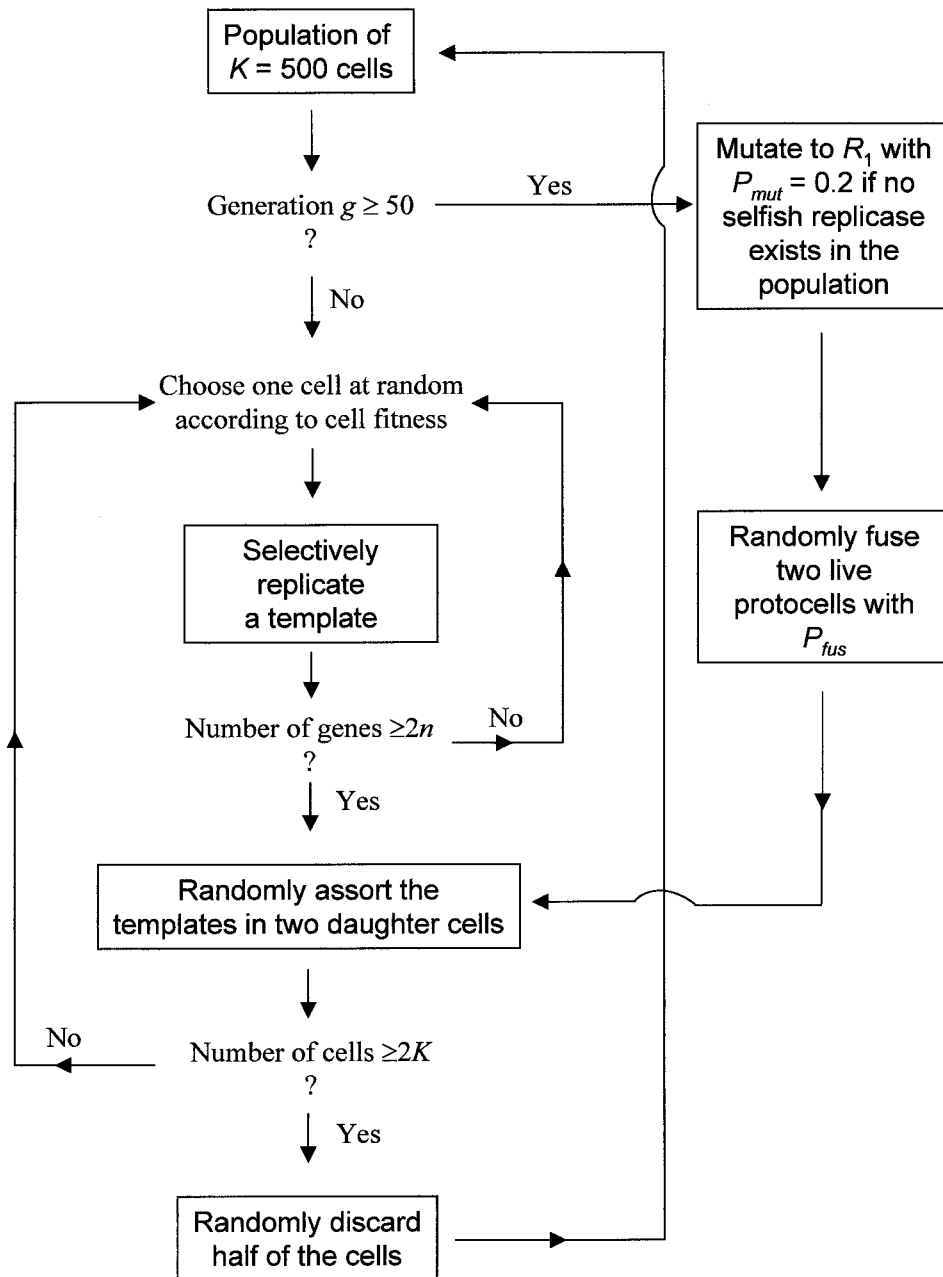
Figure 2. Flow diagram of the Monte Carlo model with no input of deleterious mutations.

Replication rate constants (μ's) are simply the products of the target affinities times the replicase activity, which here is always set to its maximum value of 1 because no deleterious mutations depressing the travelling speed of the replicase along the target template are assumed to arise. Therefore, probabilities of replication are only a function of 'target efficiencies'. Thus, we can easily model for a selfish molecule that replicates rapidly and reaps the benefits of a common metabolism, which could eventually destroy the compartment since it will increase in proportion to the expense of other genes. The reason why we have used the subscript $_j$ to denote the target replicase template in Equation (2) is because we extend the model by considering that kind of mutants.

Sex was included in the model after the population reaches its equilibrium under clonal selection (which happens well before the first 50 generations) by allowing the random fusion (with probability P_{fus} per generation) of two live protocells followed by the random re-assortment of their genomes. Potential coexistence of different replicases (i.e., selfish and/or non-selfish genes) within a given protocell is an obvious consequence of cellular fusion. When the target template is the replicase, we must distinguish between self-replication by the same kind of replicase (i.e., $j = k$ in Equation (2)), and replication by a different replicase ($j \neq k$). Depending on the relative values of self-replication rates, we have modelled for selfish (R_1; $\mu_{11(M)} < \mu_{11(R)} > \mu_{21(M)}$), non-selfish ($R_2$; $\mu_{12(M)} \approx \mu_{22(R)} \approx \mu_{22(M)}$), or altruistic ($R_3$; $\mu_{13(M)} > \mu_{33(R)} < \mu_{23(M)}$) molecules. When $j \neq k$, replication rates for target replicase were assumed to be similar to the altruistic situation; i.e., $\mu_{jk(R)} < \mu_{jj(R)}$; $j = 1, \ldots, m$; $k = 1, \ldots, m$; $j \neq k$. In other words, replicases derived from different ancestors do not help each other.

In the simulation model (Figure 2) we have considered a population of protocells initially heterogeneous for a non-selfish replicase at the compartment level. Thus, 50% of the protocells hosted a replicase R_2 and the other 50% a replicase R_3 (see above). After 50 generations of clonal selection the population reaches its steady-state relationship between selection and the random loss of any gene after stochastic assortment of templates. Then, a random protocell is chosen and all its replicase templates are 'mutated' to a selfish R_1. This is an unrealistic situation because mutation will only affect one molecule. However, the previous assumption helps to minimise the stochastic loss of the selfish replicase in the first generations after introduction and does not qualitatively change the conclusions. With probability P_{fus}, two random live protocells undergo fusion and random re-assortment of their templates. This sexual phase, where cells 'recombine' their genomes, is separate from, and precedes, template replication and reproduction. After fusion recombinant protocells resume their place in the population. If the selfish replicase is lost from the population, each successive generation has a probability of 0.2 of reintroducing a mutated protocell with R_1. The selfish replicase is only reintroduced if it is lost; the probability 0.2 merely allowing some time to elapse before a new introduction. Each simulation is continued for 1500 generations.

2.2. DELETERIOUS MUTATIONS

Here we describe the simulations to explore the combined effects of the horizontal spread of selfish genes as a consequence of sex and the continuous input of deleterious mutations that impair protocell fitness (in the realistic range in which stable mutant distributions of quasi-species within compartments are established; see Szathmáry and Demeter, 1987; Zintzaras *et al.*, 2002). The MATLAB (1999) version developed for the dynamically continuous case of the stochastic corrector model (Zintzaras *et al.*, 2002) was used after some modifications to allow protocell fusion and coexistence of different replicases within protocells. Briefly, the Monte Carlo model is the same as above (see Figure 2) but each template in Figure 1 is now assumed to consist of three mutable sites (nucleotides) with a deleterious mutation rate per nucleotide per replication round equal to u. At the protocell level the fitness function exponentially decreases from $w_{\max} = 1$ to 0 depending on the number of mutant nucleotides per metabolic gene. Replication rates are the products of the target affinities multiplied by the replicase activities, but now these rates are determined by the entries in (k, l) matrices depending on the number of deleterious mutants of the target template ($k = 0, \ldots, 3$) and the replicase ($l = 0, \ldots, 3$; see Zintzaras *et al.*, 2002). Therefore, when allowing for mutation and coexistence of different ancestral replicases the parameter space increases considerably. Here we have limited ourselves to study the relevant situation of considering wild-type selfish and non-selfish replicases as those above (i.e., R_1, R_2, and R_3) while keeping all parameters constant when target template and/or replicase have at least one deleterious mutant.

3. Simulation Results and Discussion

3.1. COMPARTMENT (CLONAL) SELECTION

As previously indicated, three qualitatively contrasted situations are conceivable in the stochastic corrector model, and their respective behaviours are plotted in Figure 3. First, a selfish replicase reaps the benefits of a common metabolism and swiftly outgrows the metabolic genes, which could eventually drive the lineage to extinction (the possible extreme of selfishness; e.g. the situation for $R_{1(a)}$ in Figure 3). Second, all metabolic and replicase templates grow at nearly similar rates, which optimises the proportion of genes within compartments and, hence, minimises the costs of the irregular transmission mechanism (R_2). Third, an altruistic replicase helps the metabolic genes but at the expense of deviating the whole compartment from the optimal gene composition (R_3). Under a clonal population structure it is clear that compartments with cooperating molecules can be stable against invasion by selfish mutants because there is ample opportunity for between-cell selection to overcome within-cell selection, and the proportion of protocells

with a non-selfish replicase would quickly increase (see also Szathmáry and De-meter, 1987; Szathmáry, 1989). We could easily assume that this 'symbiotic' group would ultimately growth in size and complexity, followed by the evolution of new structural and catalytic molecules before a more complex organism developed (De Duve, 1991; Zintzaras *et al.*, 2002).

We now digress to point out some caveats in the suggestion that linkage of genes to form a primitive chromosome could spread by selection at the level of individual replicators in the stochastic corrector model because it increases the chance for daughter protocells to have a complete set of genes (Maynard Smith and Szathmáry, 1993, 1995). Thus, if we reasonably assume that replication rates were not constant but there was instead genetic variation for target efficiencies at the population level, Figure 3 clearly shows that clonal selection is very efficient for protocells to come near the 'optimal' composition of genes. Furthermore, be-cause the probability of producing a daughter protocell lacking an essential gene rapidly approaches zero when redundancy becomes larger, the conclusion is that in a statistical sense clonal selection in the stochastic corrector model strongly favours those lineages enjoying the benefits of a 'regular-like transmission sys-tem'. Therefore, it appears that linkage could only have evolved after selection would benefit those protocells with less redundant genomes. We can hypothesize two scenarios where this might have happened. First, it could be the case that the transition from original self-replicating RNA-like molecules (ribozymes) to more efficient polymers that evolved the ability to code for proteins and uncoupled gene replication from enzyme replication (increasing the costs of redundancy) preceded the transition from independent replicators to chromosomes. Second, some form of symmetry breaking in the stoichiometric proportion of + and − strands could have opened the possibility for a very early origin of 'transcription' (Szathmáry and Maynard Smith, 1993). Replicative bias could happen if target affinities of + and − templates were different, and it would pay to make more 'enzymes' (say the + strand) than genes (the − strand). In this case, linkage of limiting minus-strands for all genes would increase the proportion of daughter cells containing a complete set of genes. It would be very interesting to explore under what set of conditions the establishment of a linkage group of genes (chromosome) is assured in the second (and probably the simplest to modelling) scenario.

3.2. 'SEX BETWEEN PROTOCELLS': NO DELETERIOUS MUTATIONS

Cellularization created a new level of selection and clearly aligned the immediate benefits of each gene with those of the whole genome, but it did not totally solve the conflict between short- and long-term evolutionary strategies. Dyson (1999) made the sensible suggestion that the first self-replicating molecules might have been 'viruses' (a very unfortunate term in this context) that could have preyed on bags of molecules and multiply. His idea falls very short to the proposal that selfish replicators somehow evolved the capacity for their transfer to other proto-
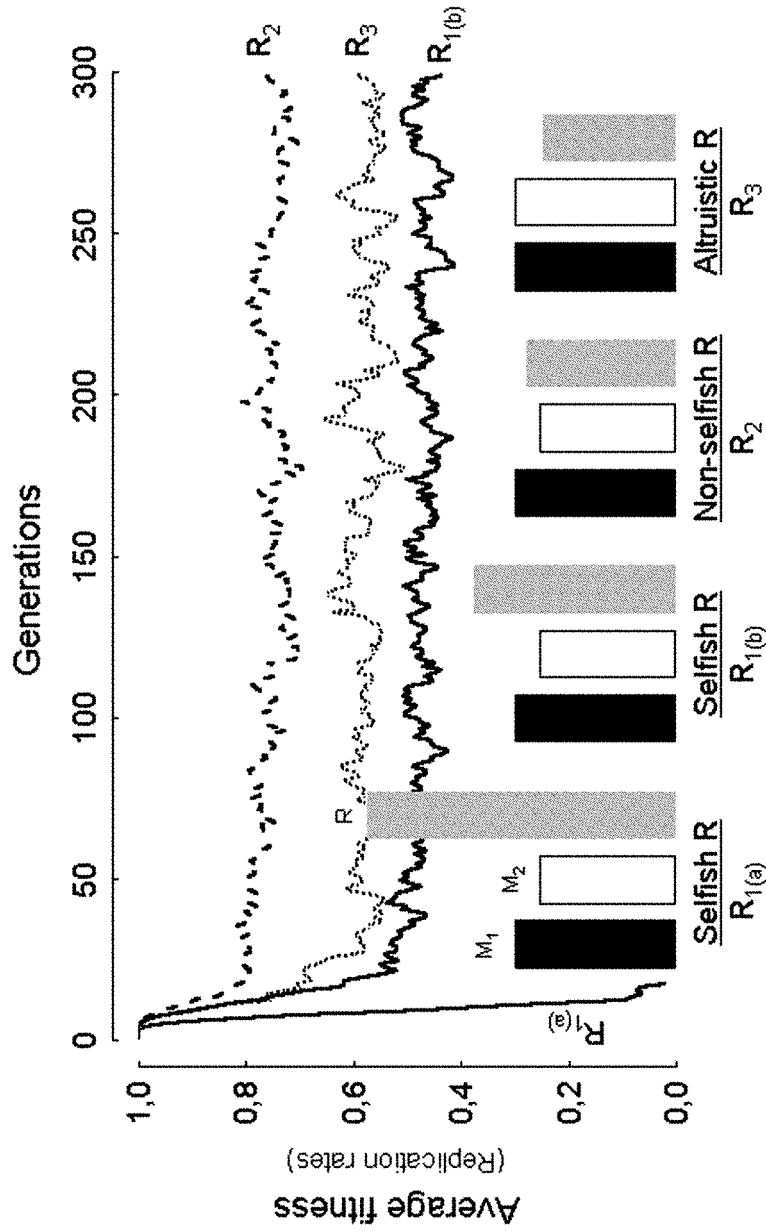
Figure 3. Trajectories for average fitness in four different sets of clonal populations according to the sort of replicase gene (R_j) hosted by protocells. Replication rates for metabolic (M_1, M_2) and replicase templates are plotted at the same scale than average fitness. At generation t_0 all protocells start with 20 copies of each template; i.e. a total of 60 genes. Twelve independent runs with 150 protocells per generation were obtained in each case, and the total average fitness is plotted. For any particular simulation we assumed that the population collapses if the average fitness of protocells was lower than 0.05. $R_{1(a)}$ is the extreme case of a selfish replicase that quickly over-exploits protocells for rapid, short-term benefits and extinction of the population occurs within the first 20 generations.

cells and became 'infectious', their evolutionary success being directly dependent on their ability for horizontal transfer. This clearly suggests primitive forms of conjugation. In addition, Redfield (2001) has presented convincing arguments that competence for oligonucleotide uptake by primitive unicellular organisms as a way of getting extra nucleotides and energy is a trophic adaptation, and Sagan and Margulis (1987) have also suggested that cannibalism by primitive cells in times of starvation could have evolved to a stalemate, with cells becoming fused but eventually separating when the environmental conditions improve.

Therefore it is thoroughly unsound to take for granted that compartment selection was operating in a natural and continuous manner as discussed in the preceding section. Lateral gene transfer between cells and/or 'accidental' uptake of selfish replicators could have been significant in protobionts, but now the question naturally arises: could a putative population of protocells resist invasion of a horizontally transmissible parasite? As stated the enquiry is obviously superfluous because we already know that life on earth has been very successful. However, apparently well-grounded models have to be tested against all possible alternatives before obtaining meaningful conclusions.

In the simulation model (Figure 2) we have considered a population of protocells initially heterogeneous for a non-selfish replicase at the compartment level. Thus, 50% of the protocells hosted replicase R_2 and the other 50% replicase R_3 (see Figure 3 for replication rates). Assume that once the clonal population reaches its steady-state relationship between selection and assortment load a single mutated protocell, which hosts a selfish replicase R_1, appears. Also assume that at the population level there is a certain probability (P_{fus}) that two protocells can randomly fuse and re-assort their genomes. The fate of R_1 will obviously depend on the manifold combinations of parameter values we could consider, such as potential number of different replicase molecules (R_j; see Figure 1) within a given protocell, replication rates, and probability of cellular fusion per generation. However, some clear conclusions emerge from extensive numerical results.

Unless the relative magnitude of cellular fusion is low, a non-lethal selfish replicase (e.g. $R_{1(b)}$ in Figure 3) would eventually reach fixation, decreasing the average fitness of the population initially homogeneous for cooperating genes (Figure 4a). On the other hand, a lethal selfish replicase can initially spread in the population only if P_{fus} is moderate or high (Figure 4b). Low fusion rates enhance the probability of successful symbiosis within protocells, and the threshold for cooperative evolution within compartments decreases with the degree of gene selfishness. This is because cell fusion must happen before protocells hosting a cooperative group of molecules outcompete those with non-cooperative groups, which is a very fast outcome in the case of a selfish replicase that quickly overexploits protocells for rapid, short-term benefits like $R_{1(a)}$ (Figure 3). Worth noting is that these behaviours are mostly independent on the ploidy level as far as the number of gene copies per cell is large enough to avoid an unsupportive assortment load. Thus, simulations with $n = 15$ templates per cell at t_0 (i.e., 5 copies of each

template) lead to the same qualitative conclusions (data not shown). Overall, the results are hardly surprising and clearly resemble the widely hypothesized fundamental conflict between horizontal (infectious) and vertical (intergenerational) modes of parasite transmission (Levin and Pimental, 1981; Anderson and May, 1982; Ewald, 1983).

An important point here, however, is that a protocell population could resist invasion of rapid exploitation by a potentially lethal (at the compartment level) parasite (Figure 4b). Although the claim of 'general principles' based on purely numerical work is always dangerous, this conclusion seems to be robust in view of the repeated appearance of more or less stable coexistence patterns of lethal and cooperative replicases (as that observed in Figure 4b) in all simulations performed with different initial conditions. Interestingly, for a given P_{fus} the likelihood for repressing the short-term success of the parasite (i.e., to decrease the number of copies of a selfish replicase per cell) is strictly dependent on the replication rates of cooperative replicases. Thus, a very simple 'suppression strategy' to favour the higher-level (protocell) units is for a group member replicase to match the replication rates of the metabolic (M_i) genes with its own self-replication rate (a form of mutual 'policing' by the replicases; see Frank, 1995).

We have previously claimed that it is wrong to suggest (Bernstein *et al.*, 1984) that cell fusion would be beneficial at the population level provided that genes that cause intragenomic conflict are taken into account. However, this may not always be the case. Thus, we have run some simulations where the ongoing process of cellular fusion (Figure 2) is stopped after a number of generations and the population turns back to clonal reproduction. In the case of a non-lethal replicase (e.g., Figure 4a) the population can get rid of the parasite and its average fitness increases. However, after invasion by an over-exploiting replicase (e.g., Figure 4b) all or most protocells host a substantial number of copies of the lethal parasite and clonal selection could be catastrophic for the entire population. In this situation extensive cellular fusion would be beneficial and the argument by Hamilton *et al.* (1990) for sex as an adaptation to parasites applies.

Perhaps a better approach in the simulations would involve the assumption that cell fusion is not entirely random but precisely driven by the selfish replicase (R_1) to promote its own survival (see Hickey and Rose, 1988); in other words, a protocell with this gene ('donor') can fuse (with a certain probability) and re-assort its genome with another randomly picked protocell ('receptor'). This could be visualized as a primitive transfer mechanism that could have mediated the whole transfer of protocell's genome as an accidental consequence, which subsequently evolved for the transfer of plasmid genes (Redfield, 2001). Computer simulations in these cases suggest that the qualitative conclusions are about the same as those previously obtained, but the threshold for cooperative evolution within compartments requires low probabilities of cellular fusion. The population of protocells could also resist invasion of a lethal selfish gene (data not shown).
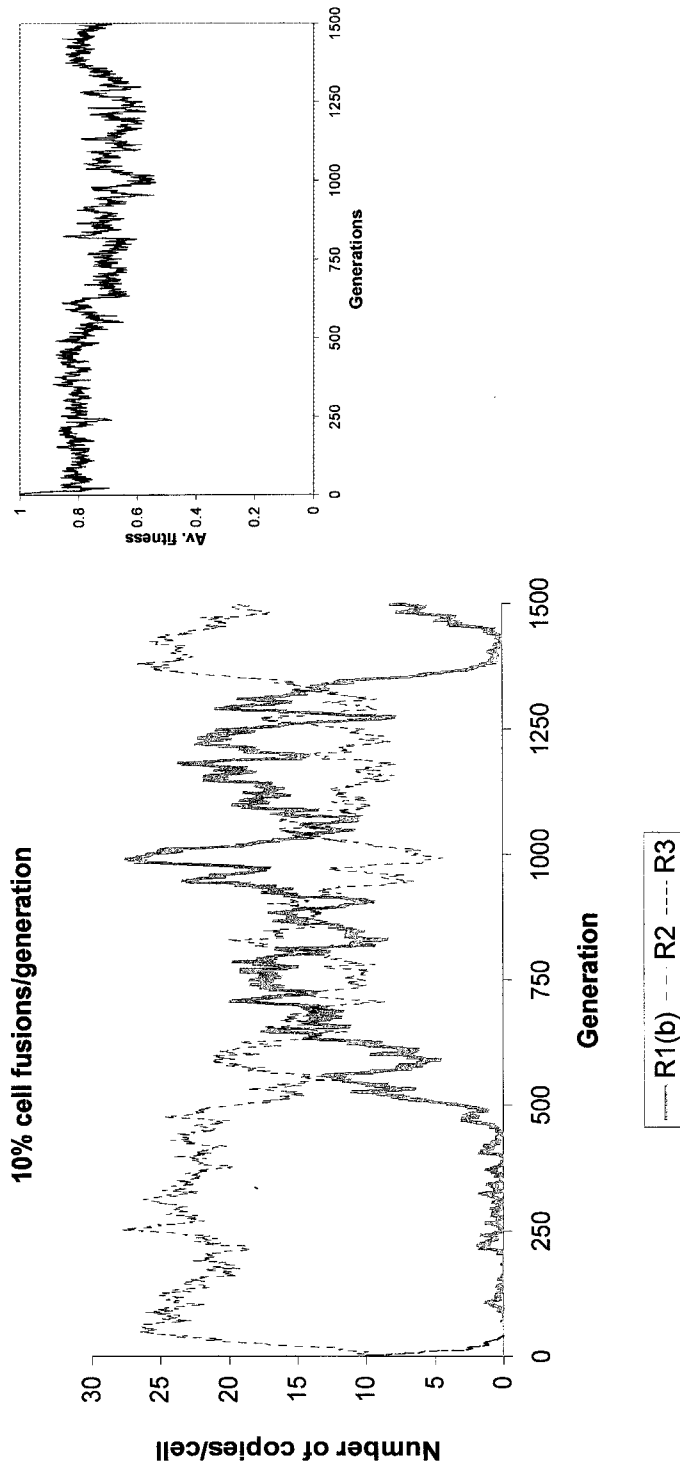
Figure 4a. Sample simulations showing the average number of replicase templates (R_j) per protocell and the average fitness of the population (embedded plots) according to the proportion of cells that undergo random cellular fusion after the 50th generation. At the compartment level, the initial population of 500 protocells with 20 copies of each gene (M_1, M_2 and R_j) was assumed to be homogeneous for a common ancestor replicase, but the proportion of protocells enclosing a non-selfish (R_2) or altruistic (R_3) replicases was about 50% in each case. After 50 generations of clonal selection (notice that the population quickly becomes homogeneous for R_2), a random protocell was chosen and all its copies for the replicase template were 'mutated' to a selfish replicase R_1. If the selfish replicase is lost, it is reintroduced with probability 0.2. (a) Non-lethal selfish replicase. The 'fusion loads'; i.e. the decrease in average fitness as a result of sex (assuming that the population could eventually turn back to clonal reproduction and get rid of the parasite; see text for details), defined as $L_{\text{fusion}} = (w_{\text{clonal}} - w_{\text{sex}})/w_{\text{clonal}}$, where w_{sex} and w_{clonal} are the average fitnesses with and without sex, respectively, are: ~0.043 (10% cell fusions/generation), ~0.158 (20%) and ~0.335 (40%).
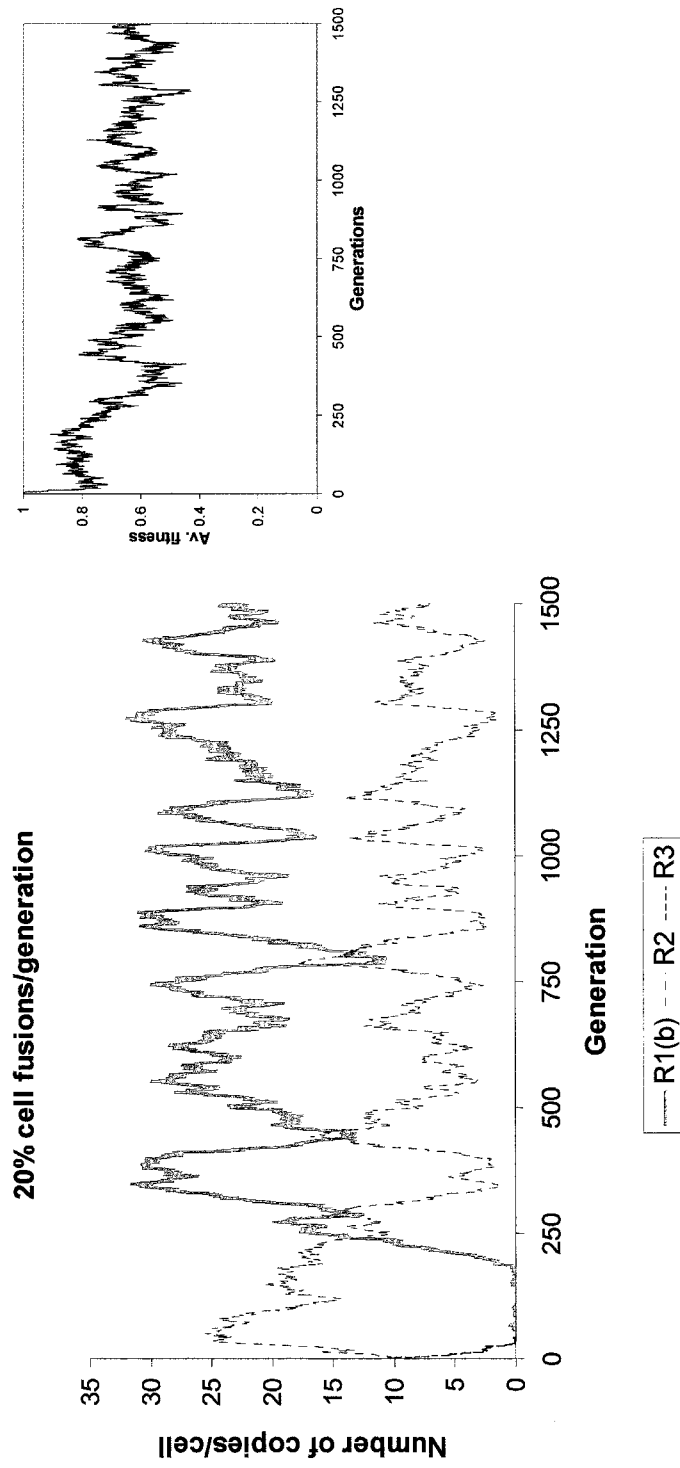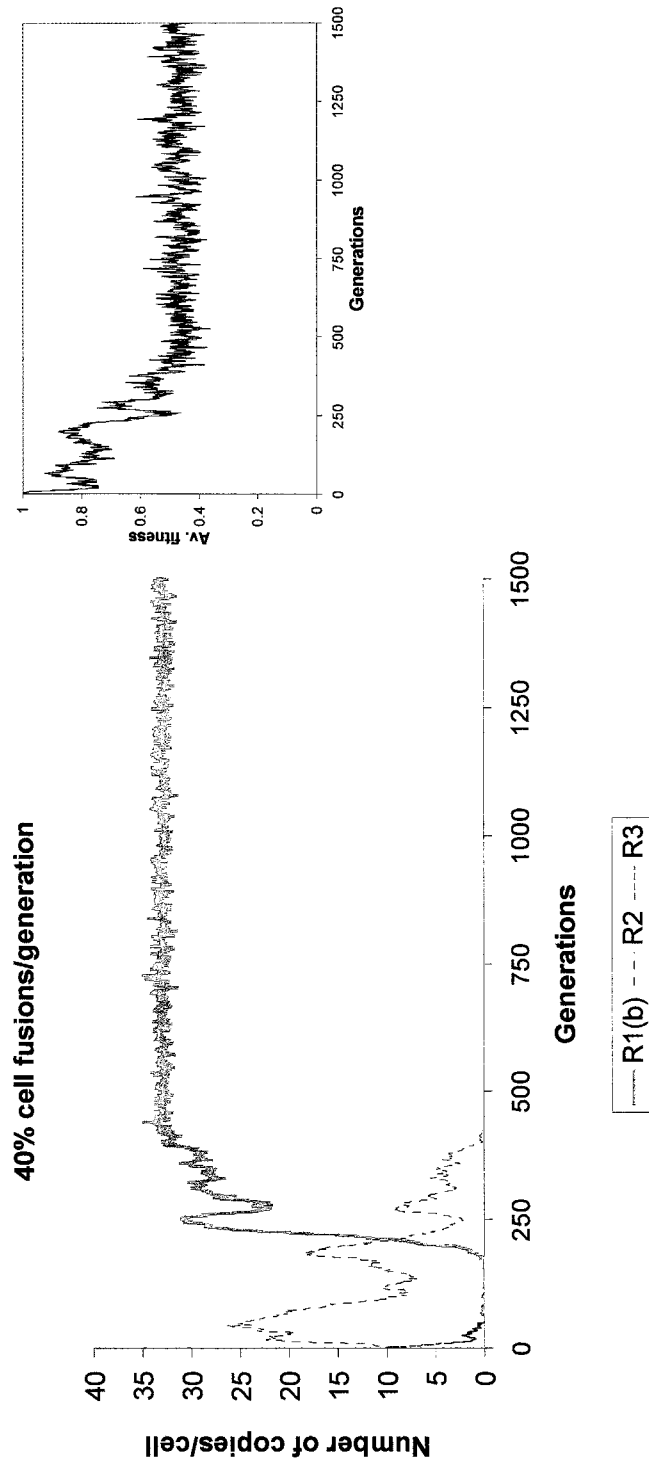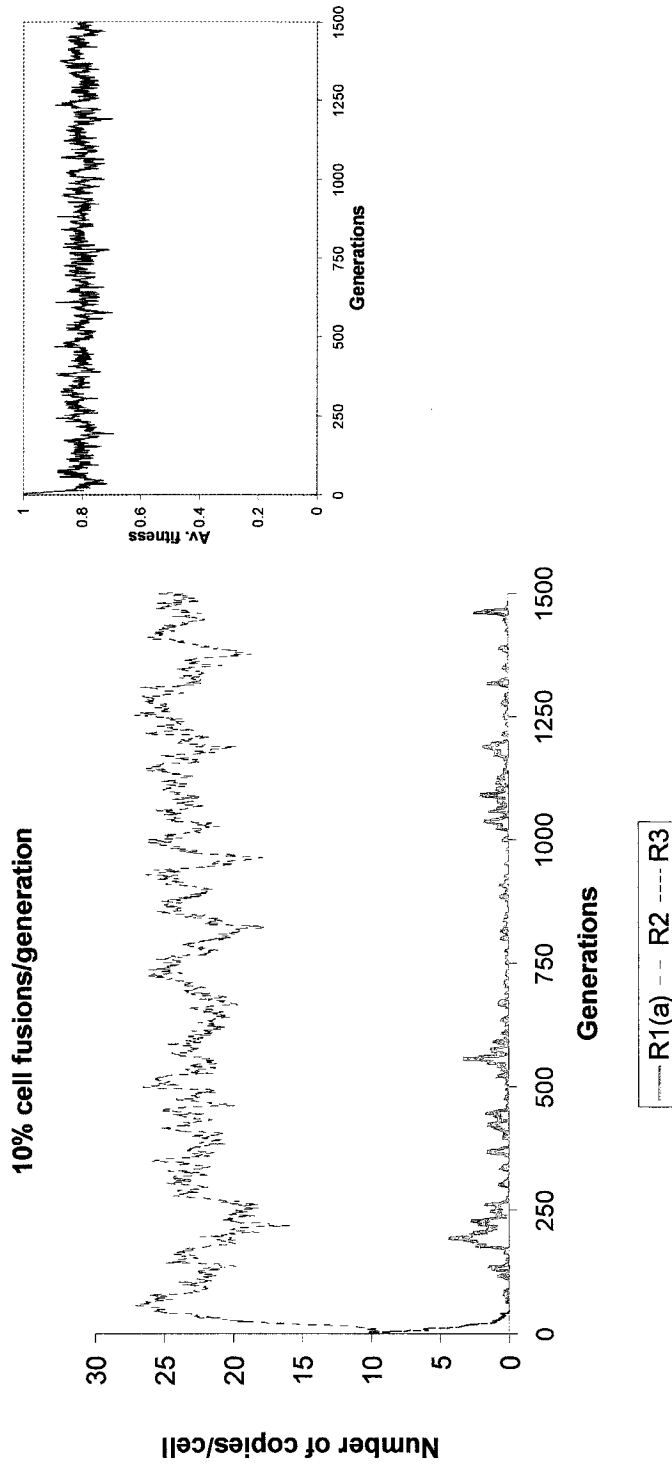
Figure 4a. (continued).

Figure 4a. (continued).

Figure 4b. Sample simulations showing the average number of replicase templates (R_j) per protocell and the average fitness of the population (embedded plots) according to the proportion of cells that undergo random cellular fusion after the 50th generation. At the compartment level, the initial population of 500 protocells with 20 copies of each gene (M_1, M_2 and R_j) was assumed to be homogeneous for a common ancestor replicase, but the proportion of protocells enclosing a non-selfish (R_2) or altruistic (R_3) replicases was about 50% in each case. After 50 generations of clonal selection (notice that the population quickly becomes homogeneous for R_2), a random protocell was chosen and all its copies for the replicase template were 'mutated' to a selfish replicase R_1. If the selfish replicase is lost, it is reintroduced with probability 0.2. (b) Lethal (at the compartment level) selfish replicase. Fusion loads: ~0 (10%), ~0.065 (20%), and ~0.167 (40%).
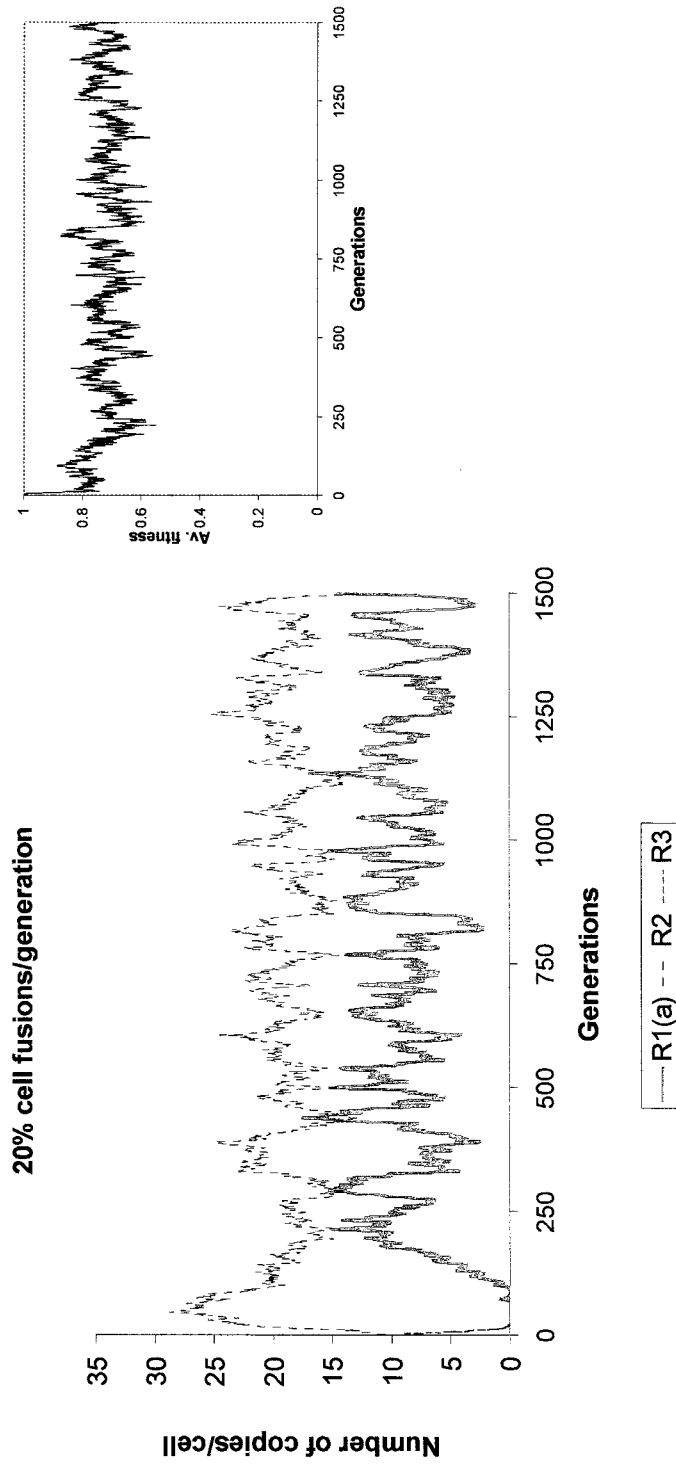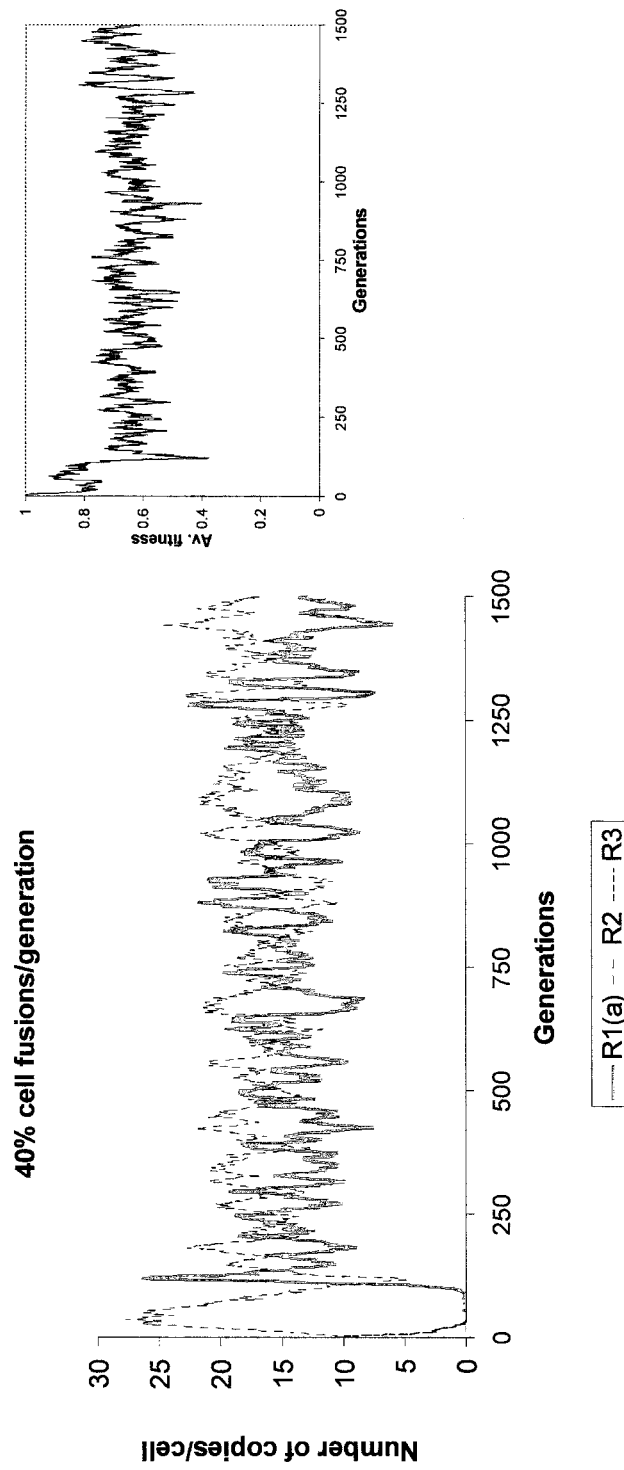
Figure 4b. (continued).

Figure 4b. (continued).

Finally, in the foregoing discussion we did not take into account the cellular-mechanistic cost of sex: the time it takes for two cells to fuse and re-assort their genomes. This cost was implemented in some further simulations by simply storing newly 'recombinant' cells for a given length of time ('dormant phase'), when these cells do not replicate their genes so that they cannot grow up. Dormant phases varied between 20–50% of population growth (i.e., only those cells that did not fuse at the start of generation t_g were allowed to grow until population size increased to 600–750 cells; see Figure 2), and a wide range of trajectories can obviously be observed because of the many parameters in the model. However, the important point here is to indicate that the prospects of population invasion by selfish genes did not dramatically change.

3.3. 'SEX BETWEEN PROTOCELLS': DELETERIOUS MUTATIONS

Most mutations are deleterious, and before the evolution of efficient replication machinery the copying fidelity per nucleotide was probably between 0.99 and 0.90 (Friedberg *et al.*, 1995; see Johnston *et al.*, 2001). At the reproducing compartment level the stochastic corrector model is formally analogous to Eigen's (1971) quasi-species model for replicating macromolecules (Szathmáry and Demeter, 1987; Zintzaras *et al.*, 2002). Darwinian selection is expected to favour those evolutionary units with high reproduction rates; however, the interplay between selection and mutation is complex when the input of deleterious mutations is very high. Thus, simulations by Wilke *et al.* (2001) indicate that slow reproducers can outcompete faster replicating counterparts at high mutation rates.

The two-level selection in the stochastic corrector model (i.e., within-compartment selection since genes are unlinked and free to compete, and between-compartment selection due to differential growth of protocells) raises important challenges on the speed of adaptive evolution under mutation-selection balance, particularly when the potential spread of selfish genes because of cellular fusion is taken into account. Figure 5a plots some sample simulations showing the average number of wild-type replicases per protocell when a non-lethal selfish replicase ($R_{1(b)}$) appears in the population. In contrast to the previous findings (Figure 4a), the selfish gene reaches fixation even when the relative magnitude of cellular fusion is low. An interesting result is that under certain circumstances the average fitness of the population slightly increases after invasion by a selfish replicase. Thus, when mutation rates are high ($u = 0.025$) and there is a large probability of cellular fusion, the average equilibrium fitness of protocells hosting the non-lethal selfish replicase can be higher than that for protocells hosting a cooperative (R_2) replicase. The reason is that protocell fitness is a decreasing function of the number of mutations in metabolic genes essential for cell growth, and higher replication rates for the replicase results in slower replication and higher concentration of less-mutated metabolic genes per protocell. On the other hand, coexistence of lethal ($R_{1(a)}$) and cooperative replicases can be observed at low mutation rates but the outcome

switches to favour the latter as mutation rate was increased (Figure 5b). These findings demonstrate the importance of mutation rates in the evolutionary fate of parasites. In addition, they suggest that under certain conditions sex might be beneficial at the protocell level by increasing the speed at which selfish replicases that raise protocell's average fitness spread in the population under mutation-selection balance, a somewhat counterintuitive outcome.

4. Conclusions

If sex is defined as the exchange of genetic material between genomes (Michod and Levin, 1988), then sexual reproduction has a long evolutionary history and is extremely widespread in nature. However, Cavalier-Smith (2002) has recently summarized his views on the origins of sex and recombination and defines as 'true' sex the combined presence of syngamy, nuclear fusion and meiosis. Under this definition sex could have only originated after the major transition from prokaryotes to eukaryotes, some 1.2 billion years ago or so (Knoll, 2003). Therefore, he clearly decouples the origin of sex from the origin of the machinery for DNA recombination, which dates circa the origin of life, and is very critical with the idea that 'parasexual' mechanisms (transduction, transformation, and plasmid conjugation) can be considered as real sex.

In our view, however, it is theoretically plausible that an evolutionarily continuous exchange of genetic material (including parasexual mechanisms) exists (see Maynard Smith and Szathmáry, 1995). By allowing the spread of the essential evolutionary units (genes) whose 'self-interest' drives natural selection, sex has promoted genomic conflicts since the very early origin of life (see Hamilton *et al*., 1990). It is unfortunate that the 'sober' question of the origin of sex quickly turns round to the question: what is the good of sex? In order to find 'adaptive function' for sex, Bernstein's *et al*. (1984) repair argument missed some fundamental problems and offered a wrong answer. Gene redundancy in primitive protocells assured a high probability of transmission of essential gene copies to daughter cells and provided a safeguard to genetic damage, and it seems very unlikely that the origin of the basic machinery for general recombination is selectively coupled to the genetic exchange between genomes (see Cavalier-Smith, 2002). Nevertheless, our simulation results suggest that as a side effect of the dynamic equilibrium between cellular fusion-mutation-selection sex could have been beneficial for primitive protocells under some conditions (see above). Were sex beneficial or detrimental to protocells, the scenario we have explored numerically is fully consistent with the idea that life may have begun as a series of ever-changing, swapping committees of proto-organisms that exchanged much genetic information (Woese, 1998).

Yet one must wonder about the mechanistic feasibility of genetic exchange among protocells. Lipid vesicles readily form under appropriate conditions, and can multiply when the necessary membrane building blocks are added internally
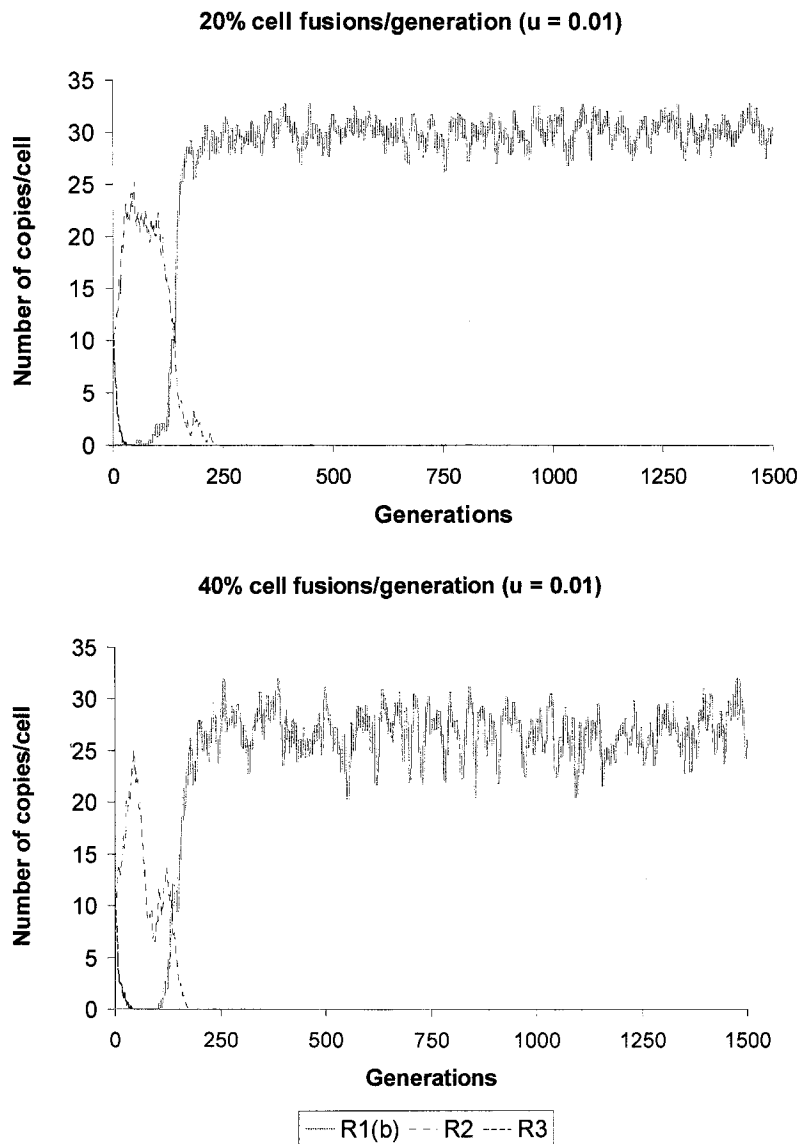
20% cell fusions/generation (u = 0.01)

40% cell fusions/generation (u = 0.01)

R1(b) — — R2 ---- R3
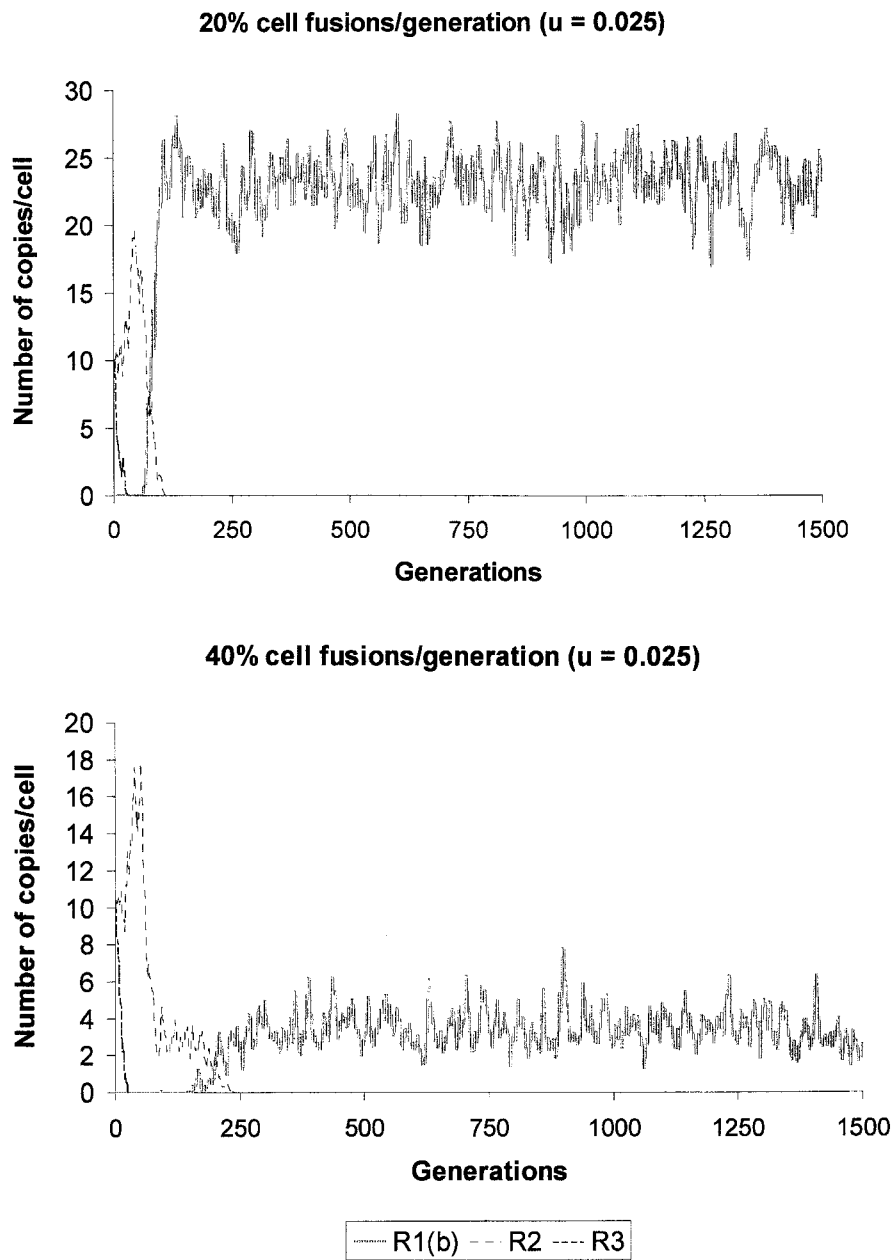
Figure 5a. Sample simulations showing the average number of wild-type replicase templates (R_j) per protocell according to the proportion of cells that undergo random cellular fusion after the 50th generation and the mutation rate u per nucleotide and replication round. At the compartment level, the initial population of 500 protocells with 20 copies of each gene $(M_1, M_2$ and $R_j)$ was assumed to be homogeneous for a common ancestor replicase, but the proportion of protocells enclosing a non-selfish (R_2) or altruistic (R_3) replicases was about 50% in each case. After 50 generations of clonal selection a random protocell was chosen and all its copies for the replicase template were 'mutated' to a selfish replicase R_1. If the selfish replicase is lost, it is reintroduced with probability 0.2. (a) Non-lethal selfish replicase. Average fitnesses at equilibrium were: 20% cell fusions/generation, $\sim$0.516 $(u = 0.01)$ and $\sim$0.498 $(u = 0.025)$; 40% cell fusions/generation, $\sim$0.503 $(u = 0.01)$ and $\sim$0.381 $(u = 0.025)$.

M. SANTOS ET AL.

20% cell fusions/generation (u = 0.025)

40% cell fusions/generation (u = 0.025)

R1(b) — — R2 ---- R3

Figure 5a. (continued).

20% cell fusions/generation (u = 0.01)
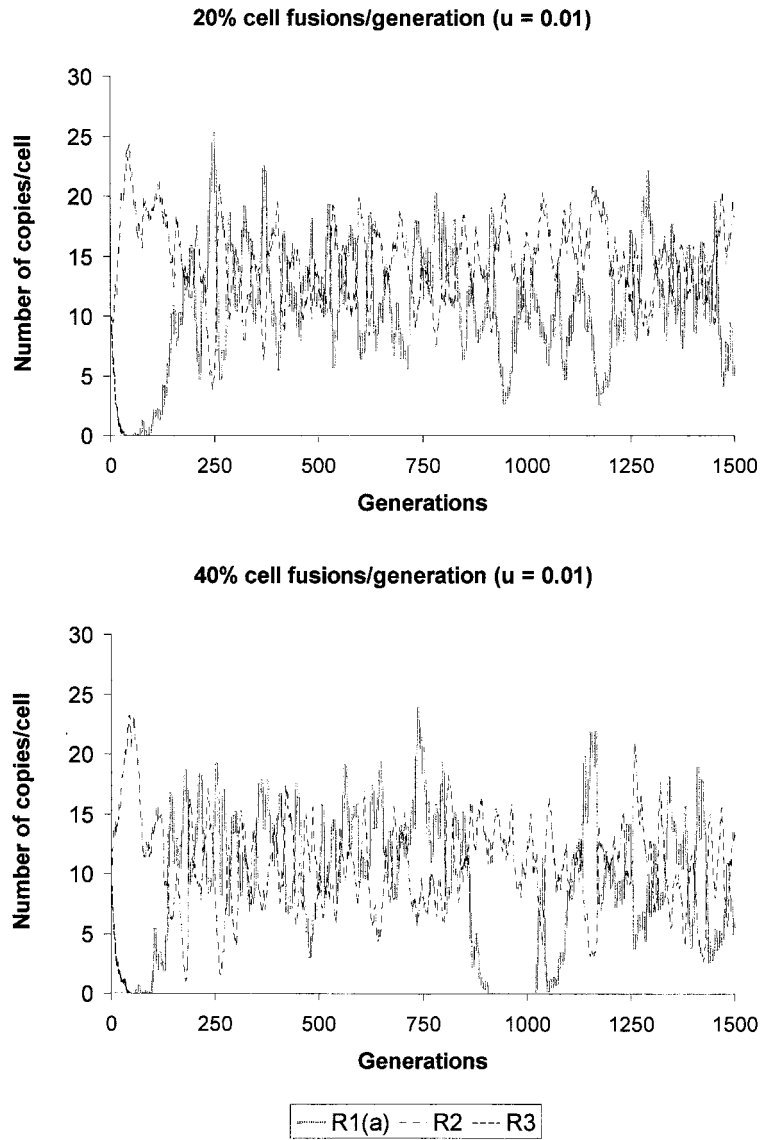

40% cell fusions/generation (u = 0.01)

R1(a) - - R2 ---- R3

Figure 5b. Sample simulations showing the average number of wild-type replicase templates (R_j) per protocell according to the proportion of cells that undergo random cellular fusion after the 50th generation and the mutation rate u per nucleotide and replication round. At the compartment level, the initial population of 500 protocells with 20 copies of each gene $(M_1, M_2$ and $R_j)$ was assumed to be homogeneous for a common ancestor replicase, but the proportion of protocells enclosing a non-selfish (R_2) or altruistic (R_3) replicases was about 50% in each case. After 50 generations of clonal selection a random protocell was chosen and all its copies for the replicase template were 'mutated' to a selfish replicase R_1. If the selfish replicase is lost, it is reintroduced with probability 0.2. (b) Lethal (at the compartment level) selfish replicase. Average fitnesses at equilibrium were: 20% cell fusions/generation, $\sim$0.644 $(u = 0.01)$ and $\sim$0.553 $(u = 0.025)$; 40% cell fusions/generation, $\sim$0.616 $(u = 0.01)$ and $\sim$0.353 $(u = 0.025)$.

20% cell fusions/generation (u = 0.025)

40% cell fusions/generation (u = 0.025)
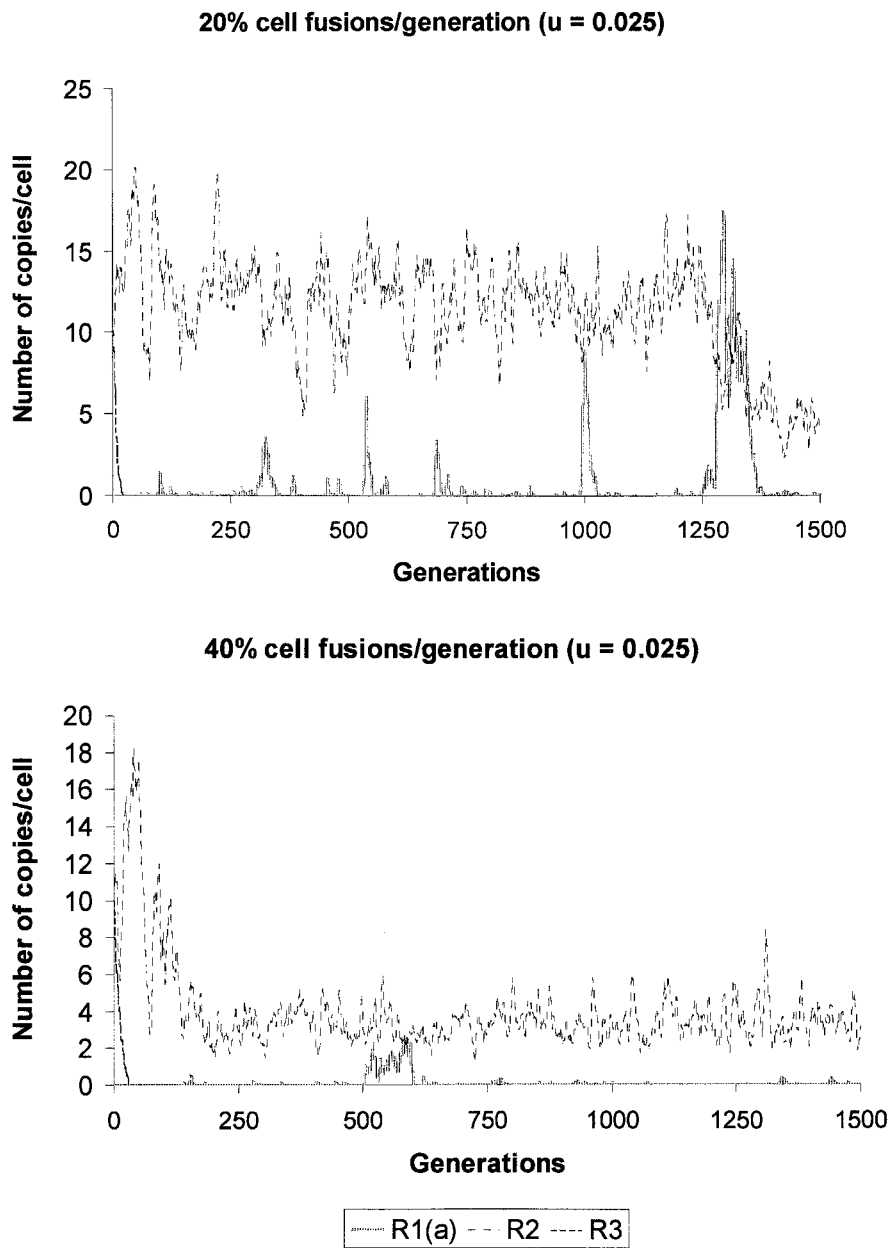

R1(a) - - R2 ---- R3

Figure 5b. (continued).

or externally (see Pohorille and Deamer, 2002 for review). Here we assume that the building blocks are generated from within, due to the metabolism of the compartment. Calculations (reviewed by Maynard Smith and Szathmáry, 1995) show that this condition can lead to vesicle growth and fission. Attempts at constructing such metabolising vesicles are under way (see Szostak *et al.*, 2001).

Fusion is a different matter, however. Analysis of liposome features (Lasic, 1998) reveals that, despite apparent difficulties, it may be a feasible option. First, vesicles of a narrow and enduring size distribution are, contrary to intuition, thought to be kinetically rather than thermodynamically controlled. This metastability can lead to accidental vesicle fusion when the vesicles critically approach one another. Or, alternatively, vesicle fusion can be triggered by a change in pH, membrane composition, or hydration. On the more active side one can imagine a transmembrane RNA/ribozyme that could couple vesicles and facilitate their fusion. Such RNA may even be analogous to a sex-inducing bacterial plasmid. When coupled to some other metabolic genes (note that linkage is not analysed in this paper), the analogy with the plasmid, hosting a battery of genes, becomes closer. There is enough room for further investigation, both experimental and theoretical.

Acknowledgements

References

Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L., Westerfield, M., Ekker, M. and Postlethwait, J. H.: 1998, Zebrafish Hox Clusters and Vertebrate Genome Evolution, *Science* **282**, 1711−1714.

Anderson, R. M. and May, R. M.: 1982, Coevolution of Hosts and Parasites, *Parasitology* **85**, 411–426.

Bernstein, H., Byerly, H. C., Hopf, F. A. and Michod, R. E.: 1984, Origin of Sex, *J. Theor. Biol.* **110**, 323–351.

Bernstein, H., Byerly, H. C., Hopf, F. A. and Michod, R. E.: 1985, Genetic Damage, Mutation, and the Evolution of Sex, *Science* **229**, 1277–1281.

Bresch, C., Niesert, U. and Harnasch, D.: 1980, Hypercycles, Parasites and Packages, *J. Theor. Biol.* **85**, 399–405.

Cavalier-Smith, T.: 1985, *The Evolution of Genome Size*, Wiley, Chichester.

Cavalier-Smith, T.: 2002, Origins of the Machinery of Recombination and Sex, *Heredity* **88**, 125–141.

Cronn, R. C., Small, R. L. and Wendel, J. F.: 1999, Duplicated Genes Evolve Independently After Polyploid Formation in Cotton, *Proc. Natl. Acad. Sci. USA* **96**, 14406–14411.

De Duve, C.: 1991, *Blueprint for a Cell: The Nature and Origin of Life*, Neil Patterson, Burlington.

Dyson, F. J.: 1999, *Origins of Life*, Cambridge University Press, Cambridge.

Eigen, M.: 1971, Self-Organization of Matter and the Evolution of Biological Macromolecules, *Naturwissenschaften* **58**, 465–523.

Eigen, M. and Schuster, P.: 1979, *The Hypercycle: A Principle of Natural Self-Organization*, Springer-Verlag, Berlin.

Eigen, M., Schuster, P., Gardiner, W. and Winkler-Oswatitsch, R.: 1981, The Origin of Genetic Information, *Sci. Am.* **244**, 78–94.

Ewald, P. W.: 1983, Host-Parasite Relations, Vectors, and the Evolution of Disease Severity, *Ann. Rev. Ecol. Syst.* **14**, 465–485.

Frank, S. A.: 1995, Mutual Policing and Repression of Competition in the Evolution of Cooperative groups, *Nature* **377**, 520–522.

Friedberg, E. C., Walker, G. C. and Siede, W.: 1995, *DNA Repair and Mutagenesis*, ASM Press, Washington.

Gánti, T.: 1987, *The Principle of Life*, OMIKK, Budapest.

Gánti, T.: 2003, *The Principles of Life*, Oxford Univ. Press, Oxford.

Gesteland, R. F., Cech, T. R. and Atkins, J. F. (eds): 1999, *The RNA World*, 2nd ed., Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.

Gilbert, W.: 1986, The RNA World, *Nature* **319**, 618.

Grey, D., Hutson, V. and Szathmáry, E.: 1995, A Re-Examination of the Stochastic Corrector Model, *Proc. R. Soc. Lond.* B **262**, 29–35.

Hamilton, W. D., Axelrod, R. and Tanese, R.: 1990, Sexual Reproduction as an Adaptation to Resist Parasites (A Review), *Proc. Natl. Acad. Sci. USA* **87**, 3566–3573.

Hickey, D. A. and Rose, M. R.: 1988, The Role of Gene Transfer in the Evolution of Eukaryotic Sex in Michod, R. E. and Levin, B. R. (eds.), *The Evolution of Sex: An Examination of Current Ideas*, Sinauer Associates, Sunderland, pp. 161–175.

Hurst, L. and Peck, J. R.: 1996, Recent Advances in Understanding the Evolution and Maintenance of Sex, *Trends Ecol. Evol.* **11**, 46–52.

Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. and Bartel, D. P.: 2001, RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension, *Science* **292**, 1319–1325.

Joyce, G. F., Schwartz, A. W., Miller, S. L. and Orgel, L. E.: 1987, The Case for an Ancestral Genetic System Involving Simple Analogues of the Nucleotides, *Proc. Natl. Acad. Sci. USA* **84**, 4398–4402.

Keightley, P. D. and Eyre-Walker, A.: 2000, Deleterious Mutations and the Evolution of Sex, *Science* **290**, 331–333.

Knoll, A. H.: 2003, *Life on a Young Planet. The First Three Billion Years of Evolution on Earth*, Princeton Univ. Press, Princeton.

Koch, A. L.: 1984, Evolution vs the Number of Gene Copies per Primitive Cell. *J. Mol. Biol.* **20**, 71–76.

Kondrashov, A. S. and Crow, J. F.: 1991, Haploid or Diploidy: Which is Better? *Nature* **351**, 314–315.

Kondrashov F. A., Rogozin, I. B., Wolf, Y. I. and Koonin E. V.: 2002, Selection in the Evolution of Gene Duplications, *Genome Biol.* **3**, research0008.1-0008.9 (URL: http://genomebiology.com/2002/3/2/research/0008).

Lasic, D. D.: 1998, Novel Applications of Liposomes, *Trends Biotech.* **16**, 307–321.

Levin, S. A. and Pimental, D.: 1981, Selection of Intermediate Rates of Increase in Parasite-Host Systems, *Am. Nat.* **117**, 308–315.

Lupski, J. R., Roth, J. R. and Weinstock, G. M.: 1996, Chromosomal Duplications in Bacteria, Fruit Flies, and Humans, *Am. J. Hum. Genet.* **58**, 21–27.

Lynch, M. and Conery, J. S.: 2000, The Evolutionary Fate and Consequences of Duplicate Genes, *Science* **290**, 1151–1155.

MATLAB V.5.3.: 1999, *The Language of Technical Computing*, The Math Works, Inc. URL: http://www.mathworks.com.

Maynard Smith, J.: 1978, *The Evolution of Sex*, Cambridge University Press, Cambridge.

Maynard Smith, J.: 1979, Hypercycles and the Origin of Life, *Nature* **280**, 445–446.

Maynard Smith, J. and Szathmáry, E.: 1993, The Origin of Chromosomes I. Selection for Linkage, *J. Theor. Biol.* **164**, 437–446.

Maynard Smith, J. and Szathmáry, E.: 1995, *The Major Transitions in Evolution*, W. H. Freeman, Oxford.

McDermid, H. E., and Morrow, B. E.: 2002, Genomic Disorders on 22q11, *Am. J. Hum. Genet.* **70**, 1077–1088.

Mergenthaler, S., Sharp, A., Ranke, M. B., Kalscheuer, V. M., Wollmann, H. A. and Eggermann, T.: 2001, Gene Dosage Analysis in Silver-Russell Syndrome: Use of Quantitative Competitive PCR and Dual-Color FISH to Estimate the Frequency of Duplications in 7p11.2-p13, *Genet. Test.* **5**, 261–266.

Michod, R. E.: 1993, Genetic Error, Sex, and Diploidy, *J. Heredity* **84**, 360–371.

Michod, R. E. and Levin, B. R. (eds): 1988, *The Evolution of Sex: An Examination of Current Ideas*, Sinauer Associates, Sunderland.

Niesert, U., Harnasch, D. and Bresch, C.: 1981, Origin of Life: Between Scylla and Charybdis, *J. Molec. Evol.* **17**, 348–353.

Otto, S. P. and Whitton, J.: 2000, Polyploid Incidence and Evolution, *Annu. Rev. Genet.* **34**, 401–437.

Otto, S. P. and Yong, P.: 2002, The Evolution of Gene Duplicates, *Adv. Genet.* **46**, 451–483.

Partridge, L. and Hurst, L. D.: 1998, Sex and Conflict, *Science* **281**, 2003–2008.

Perrot, V., Richerd, S. and Valéro, M,: 1991, Transition from Haploidy to Diploidy, *Nature* **351**, 315–317.

Pohorille, A. and Deamer, D.: 2002, Artificial Cells: Prospects for Biotechnology, *Trends Biotech.* **20**, 123–128.

Reanney, D. C.: 1987, Genetic Error and Genome Design, *Cold. Spring. Harb. Symp. Quant. Biol.* **52**, 751–757.

Redfield, R. J.: 2001, Do Bacteria Have Sex? *Nat. Rev. Genet.* **2**, 634–639.

Sagan, D. and Margulis, L.: 1987, Cannibal's Relief: The Origins of Sex, *New Sci.* **115**, 36–40.

Szathmáry, E.: 1989, The Integration of the Earliest Genetic Information, *Trends Ecol. Evol.* **4**, 200–204.

Szathmáry, E. and Demeter, L.: 1987, Group Selection of Early Replicators and the Origin of Life, *J. Theor. Biol.* **128**, 463–486.

Szathmáry, E. and Maynard Smith, J.: 1993, The Evolution of Chromosomes II. Molecular Mechanisms, *J. Theor. Biol.* **164**, 447–454.

Szostak, J. W., Bartel, D. P. and Luisi, P. L.: 2001, Synthesizing Life, *Nature* **409**, 387–390.

Werren, J. H., Nur, U. and Wu, C.-I.: 1988, Selfish Genetic Elements, *Trends Ecol. Evol.* **3**, 297–302.

West, S. A., Lively, C. M. and Read, A. F.: 1999, A Pluralist Approach to Sex and Recombination, *J. Evol. Biol.* **12**, 1003–1012.

Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E. and Adami, C.: 2001, Evolution of Digital
    Organisms at High Mutation Rates Leads to Survival of the Flattest, *Nature* **412**, 331–333.
Williams, G. C.: 1975, *Sex and Evolution*, Princeton University Press, Princeton.
Woese, C.: 1998, The Universal Ancestor, *Proc. Natl. Acad. Sci. USA* **95**, 6854–6859.
Zintzaras, E., Santos, M. and Szathmáry, E.: 2002, 'Living' under the Challenge of Information
    Decay: The Stochastic Corrector Model *versus* Hypercycles. *J. Theor. Biol.* **217**, 167–181.