An Analysis of the Hotspot Diffusion Paradox

Jérémy Barbay,
University of British Columbia,
jeremy@cs.ubc.ca

Abstract

Crossover is believed to initiate at specific sites called hotspots, by combinational-repair mechanism in which the initiating hotspot is replaced by a copy of its homologue. Boulton *et al.* studied through simulation the effect of this mechanism, and observed in their model that active hotspot alleles are rapidly replaced by inactive alleles. This is paradoxical because active hotspots alleles do not disappear in natural systems. We give a theoretical analysis of this model, which confirms their experimental result, and we argue that they failed to take properly into account the benefits of recombination, because of the optimality of their initial population. On the other hand, we show that even with an initial population of low fitness the model does not sustain the active hotspot alleles. Those results suggest that at least one model is wrong, either the one for the recombination of chromosomes, or the one for the diffusion of the hotspot alleles: we suggest another model for the diffusion of hotspots alleles.
Keywords: yeast, sexual reproduction, crossover, hotspots.

1 Introduction

When diploid organisms produce gametes, a cell reproduces and divides itself into 4 haploid gametes [3]. During this division the paired chromosomes can recombine their genes: it happens that the structure of one of the chromosomes breaks in some specific points called "hotspots". The structure is then repaired by copying the corresponding allele from the homologue chromosome, which results eventually in the exchange of whole sections of the chromosomes, phenomenon called "crossover". The hotspot alleles can be inactivated by mutations, which change them so that they do not break, and do not initiate a crossover any more. If a recombination is still initiated by the homologous chromosome, its hotspot is repaired using the mutated allele, and the gametes will contain an inactivated hotspot allele.

Boulton, Myers and Redfield [1] observe that this mechanism seems to favour the diffusion of inactive hotspot alleles, which is paradoxical since active hotspots are predominant in natural systems. They consider the positive effect of the chromosome recombination on their migration during the cell division: without recombination some gametes are sterile with probability $.5$. Their simulations show that this effect is not strong enough to sustain a positive proportion of active hotspot alleles, and they conclude that the crossover model might be wrong, as it is not self-sustainable.

In this paper we formalize Boulton *et al.*'s model (section 2), and give a theoretical analysis of it (section 3), which confirms the results of their simulations. We outline another contribution of active hotspot alleles, concerning the evolution of the species. This contribution is only feebly taken into account by the simulations of Boulton *et al.*, where the population is initially optimal. We show that even with an initial population of low fitness, the model does not sustain any proportion of active hotspots (section 4). In our conclusion (section 5) we discuss the importance of spatiality, a property lacked by the models discussed.

2 Model

We consider a pool of freely interacting haploids. Each haploid consists of a single chromosome, with a single hotspot, and is noted H_i, of fitness f_i, which corresponds to its ability to grab resources. The population size is variable, regulated by the amount N of resources (e.g. food) available. At each step of the process: with probability $.5$ one haploid H_i is chosen uniformly at random, of fit-

[1]TR-2004-12, Department of Computer Science, University of British Columbia, July 27, 2004

ness f_i, and duplicates with probability $\frac{f_i}{\sum f_i}\frac{N}{2}$, otherwise dies. Otherwise, with the remaining probability $.5$, two haploids H_i and H_j are chosen uniformly at random, of fitness f_i and f_j. They combine and try to reproduce with probability $\frac{f_i+f_j}{\sum f_i}\frac{N}{4}$, and otherwise die. If both hotspot alleles are inactive, the haploids are inviable with probability $.5$, otherwise they are duplicates of the original haploids.

If both hotspot alleles are active, the haploids combine by a crossover at this site, and the four haploids produced have active hotspot alleles. If only one hotspot allele is active, the chromosomes combine by a crossover at this site, but three out of four haploids produced have inactive hotspot alleles, and the one with an active hotspot is a duplicate of the original chromosome whose hotspot allele was active.

Note that for simplicity, the active hotspot alleles *always* break: in a model with only one hotspot allele this doesn't change the result and simplifies the proofs.

Lemma 1 *The average population size converges to N.*

3 Segregation effect

The segregation produces aneuploids (inviable gametes) with higher probability when no crossover occurs. As this happens more often when there are fewer active hotspot alleles, Boulton *et al.* studied the effects of this property. Their simulations show that this force is not sufficient to maintain a positive proportion of active hotspots alleles. They use a model with a single chromosome per individual, and a single hotspot per chromosome. We give the theoretical analysis of a broader class of models, which shows that the active hotspot alleles indeed disappear in the stationary distribution.

Theorem 1 *Starting from an optimal population, active hotspot alleles disappear in the stationary distribution.*

4 Selection effect

The active hotspot alleles also have an evolutionary utility: without them, the population evolve only by mutation. A sub-population with active hotspot alleles should have an evolutionary advantage.

Boulton *et al.*'s simulation is starting with a population of optimal individuals, and mutations introduce sub-optimal individuals. In such a model, with a small mutation rate, only a few sub-optimal individuals are generated at each generation, and there is no need of recombination to obtain a better population, as selection just suppresses sub-optimal individuals. On the other hand, with a mutation rate large enough to disrupt the optimality of the population, the possible benefits from recombination are most likely disrupted by a mutation.

With a sub-optimal initial population, recombinations are much more likely to generate better individuals, and the active hotspot alleles triggering these recombinations are more likely to be promoted. To obtain an upper bound on the proportion of active hotspot alleles sustained in such a model, it is sufficient to study a much simpler model where the offsprings obtained by recombination are always better than their parents: this model's unrealism can only increase the proportion of active hotspot alleles. In such a model, the proportion of active hotspot alleles in the stationary distribution is still null: hence even the evolutionary role of the active hotspot alleles does not permit to sustain them in Boulton *et al.*'s model.

Theorem 2 *Even in the model where new individuals are always better than their parents, the active hotspot alleles do not persist.*

5 Conclusions

In this paper we provided a theoretical confirmation of some experimental results on a model of the diffusion of active hotspots, and we gave an analysis proving that even a more general model does not solve the paradox observed.

The models discussed in the literature neglect many properties of natural systems, among which one is the spatiality. The models neglecting this property are analogous to the situation where the haploids are continuously shuffled: this happens only rarely in nature, where the haploids can form colonies.

References

[1] A. Boulton, R. S. Myers, and R. J. Redfield. The hotspot conversion paradox and the evolution of mei-

otic recombination. In *Proc. Natl. Acad. Sci. USA*, volume 94, pages 8058–8063, July 1997.

[2] G. Fayolle, V. Malyshev, and M. Menshikov. *Topics in the constructive theory of countable Markov chains*. Cambridge [England]; New York, NY : Cambridge University Press, 1995.

[3] A. W. Murray and J. W. Szostak. *Annual Review of Cell Biology*, 1:189–315, 1985.

Appendix

A Model

In this model the size S of the population is stable on average. At each step of the process, the probability that the haploids chosen duplicate and increase the population is proportional to the share of resources they get. This share is proportional to their relative fitness, but inversely proportional to the size of the population. Hence the size of the population decreases when it is too large, and increases when it is too small: the lemma 1 proves this formally.

Proof of Lemma 1:

First consider the case when a single haploid is chosen for duplication. Let be S the size of the population, ΔS the variation of this size in the first phase of one step, $E(\Delta S|\text{clone})$ the average variation when a single haploid is chosen for cloning, and $E(\Delta S|H_i)$ the average variation when the haploid H_i has been chosen:

$$\begin{aligned}
E(\Delta S|H_i) &= +1\left[\frac{f_i}{\sum_l f_l}\frac{N}{2}\right] - 1\left[1-\frac{f_i}{\sum_l f_l}\frac{N}{2}\right]\\
&= \frac{f_i}{\sum_l f_l}N - 1\\
E(\Delta S|\text{clone}) &= \sum_i E(\Delta S|H_i)\frac{1}{S}\\
&= \frac{N}{S} - 1
\end{aligned}$$

Note that ΔS is equal to 1 or -1, and that $E(\Delta S|\text{clone})$ is positive if and only if S is smaller than N.

Now consider the case when two haploids are chosen for breeding. Let be ΔS the average variation of the size of the population in this phase, $E(\Delta S|\text{breed})$ the average variation when a couple of haploids is chosen for breeding, and $\Delta S|H_i, H_j$ the average variation when the haploids H_i and H_j have been chosen:

$$\begin{aligned}
E(\Delta S|H_i,H_j) &= +2\left[\frac{f_i+f_j}{\sum_l f_l}\frac{N}{4}\right]\\
&\quad -2\left[1-\frac{f_i+f_j}{\sum_l f_l}\frac{N}{4}\right]\\
&= \frac{f_i+f_j}{\sum_l f_l}N - 2\\
E(\Delta S|\text{breed}) &= \sum_{i,j,i\neq j} E(\Delta S|H_i,H_j)\frac{1}{S(S-1)}\\
&= \frac{\sum_{i,j}(f_i+f_j) - 2\sum_i f_i}{\sum_l f_l}\frac{N}{S(S-1)} - 2\\
&= (2S-2)\frac{\sum_i f_i}{\sum_l f_l}\frac{N}{S(S-1)} - 2\\
&= 2\left(\frac{N}{S} - 1\right)
\end{aligned}$$

Note that here ΔS is equal to 2 or -2, and $E(\Delta S|\text{breed})$ is positive if and only if S is smaller than N.

Considering the two cases, the average variation of the size of the population is positive if and only if S is smaller than N. On the other hand this variation is always bounded. By analogy with a Markov chain describing ΔS, results from [2] permit to conclude that $E(S)=N$ in the stationary distribution. □

B Segregation Effect

Proof of Theorem 1: As the fitness is uniform, the duplication of haploids doesn't modify on average the repartition of active hotspot alleles, only the sexual reproduction does. Let be p the proportion of active hotspot alleles in the population. As a pair of haploids with nonactive hotspots can product 4 haploids or nothing, 3 types of pairs can be formed with 4 possible outcomes:

1. $[AA \rightarrow AAAA]$ if the two chosen haploids have active hotspot alleles;
2. $[AN \rightarrow ANNN]$ if exactly one chosen haploid has an active hotspot allele;

3. $[NN \rightarrow NNNN]$ if none of the chosen haploids has an active hotspot allele, and the gametes properly segregate;

4. $[NN \rightarrow \emptyset]$ if none of the chosen haploids have an active hotspot allele, and the gametes does not properly segregate.

In each case we study the variation in the number of active hotspot alleles and in the size of the population, expressed by the variation in the proportion of active hotspot alleles.

1. The first case happens with probability $[p^2]$, and the proportion p then increases by $\frac{4-4p}{S+4}$.
2. The second case happens with probability $[2p(1-p)]$, and the proportion p then increases by $\frac{1-4p}{S+4}$.
3. The third case happens with probability $[(1-p)^2\frac{1}{2}]$, and the proportion p then increases by $\frac{4p}{S+4}$.
4. The fourth case happens with probability $[(1-p)^2\frac{1}{2}]$, and the proportion then does not change.

The expression of the average variation of p is then expressed as the sum of the variations in each case, weighted by their probabilities:

$$\begin{aligned} E(\Delta p) &= \frac{1}{S+2}(2-2p)[p^2] \\ &\quad +(-4p)\left[2p(1-p)+(1-p)^2\frac{1}{2}\right] \\ &\quad +\frac{1}{S+2}p(1-p)^2 \\ &= \frac{1}{S+2}p(1-p)(-2)(2p+1) \\ &\quad +\frac{1}{S+2}p(1-p)^2 \end{aligned}$$

As long as $p \in (0,1)$ this is positive if and only if

$$\begin{aligned} & -2(2p+1)(S-2)+(1-p)(S+2) > 0 \\ \Leftrightarrow\quad & 6+6p-S-5pS > 0 \\ \Leftrightarrow\quad & p < \frac{S-6}{6-5S} \end{aligned}$$

For any value of S larger than 6, the fraction $\frac{S-6}{6-5S}$ is negative. This condition is fulfilled whenever N is sufficiently large (see Lemma 1).

So as long as N is large enough, $E(\Delta p)$ is never positive and p is always decreasing to 0 on average. □

C Selection Effect

Proof of Theorem 2:

As before, p is the proportion of active hotspot alleles in the population. For a random individual chosen in the population, let be A the event that it has an active hotspot allele, N the event that it has an inactive hotspot allele, and new the event that this individual is different from its parents.

Haploids with a new combination of alleles (event new) and active hotspot alleles (event A) can be generated only by breeding two haploids which both have active hotspot alleles. Haploids with a new combination of alleles (event new) and inactive hotspot alleles (event N) can be generated only by breeding one haploid, which has an active hotspot allele, with an haploid which has an inactive hotspot allele. In each case, two such haploids are generated.

From the probabilities of those events, we can deduce the probability of generating a new haploid, and the probability that such a haploid has an active hotspot allele:

$$\begin{aligned} \Pr\{A \wedge \text{new}\} &= \frac{p^2}{2} \\ \Pr\{N \wedge \text{new}\} &= \frac{2p(1-p)}{2} \\ \Pr\{\text{new}\} &= p(1-\frac{p}{2}) \\ \Pr\{A|\text{new}\} &= \frac{\Pr\{A \wedge \text{new}\}}{\Pr\{\text{new}\}} \\ &= \frac{2}{2-p} - 1 \end{aligned}$$

This number, which corresponds to the probability that a new haploid has an active hotspot allele, is smaller than p, the proportion of individuals with an active hotspot in the whole population. $\forall p \in [0,1]$:

$$\begin{aligned} \frac{2}{2-p} - 1 < p \quad & \Leftrightarrow \quad 2 < (p+1)(2-p) \\ & \Leftrightarrow \quad p(1-p) < 0 \\ & \Leftrightarrow \quad p \in [0,1] \end{aligned}$$

Hence the proportion of active hotspot alleles among new (and potentially better) solutions is *smaller* than the proportion of active hotspot alleles among the old solutions. So the active hotspot alleles do not persist. □

D More on Spatiality

A theoretical analysis shows that in one dimension, spatiality "saves" the active hotspots by permitting a colony of haploids with active hotspots to compete with its neighbouring colonies of haploids with inactive hotspots, as long as active hotspots permit to produce better solutions.

We failed to perform a theoretical analysis of the spatial model in more than one dimension, but simulations in two and three dimensions show a similar behaviour: for a finely tuned parameterization of the model, a subpopulation of haploids with active hotspots survive and compete efficiently against an environment of haploids with inactive hotspots.

One obvious perspective from here is to study spatial models of the diffusion of active hotspots. Spatial models are random cellular automata, and it is known that some cellular automata are Turing equivalent, and hence cannot be analyzed in a complete way: it is possible that a theoretical analysis of spatial models is not possible.

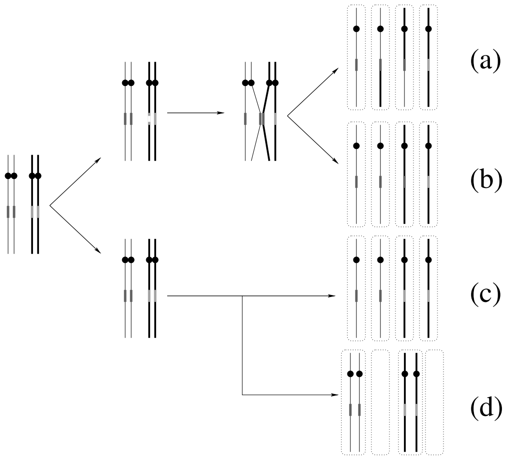

Figure 1: Crossover Mechanism: when an hotspot allele "breaks", it is repaired and replaced by the hotspot allele from the homologous chromosome, initiating a crossover (a), or not (b); when no hotspot allele break, the chromosome are just cloned (c), and sometime do not segregate properly (d).

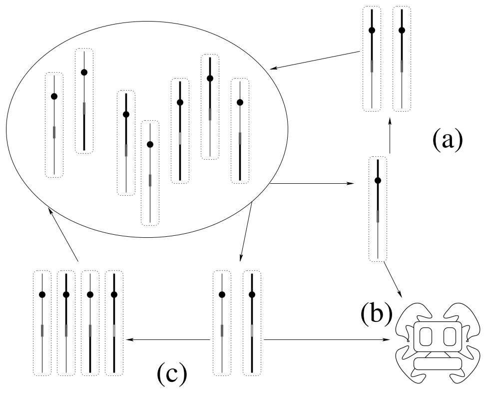

Figure 2: Life cycle of the haploids: when an individual is selected, it duplicates (a) or dies (b); when a couple is selected, its members combine (c) or die (b).